(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 February 2003 (13.02.2003)

PCT

(10) International Publication Number
WO 03/012052 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/24226

(22) International Filing Date: 30 July 2002 (30.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/308,640	30 July 2001 (30.07.2001)	US
60/370,970	8 April 2002 (08.04.2002)	US

(71) Applicants *(for all designated States except US)*: **THE GOVERNMENT OF THE UNITED STATES OF AMERICA AS REPRESENTED BY THE SECRETARY OF THE DEPARTMENT OF HEALTH AND HUMAN SERVICES** [US/US]; The National Institutes of Health, Office of Technology Transfer, 6011 Executive Boulevard, Suite 325, Rockville, MD 20852-3804 (US). **CARNEGIE INSTITUTE OF WASHINGTON** [US/US]; 115 West University Parkway, Baltimore, MD 21210 (US). **IMPERIAL COLLEGE INNOVATUIONS LTD.** [GB/GB]; Sherfield Building, Exhibition Road, London SW7 2AZ (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CAPLEN, Natasha, J.** [US/US]; 111 Normandy Drive, Silver Spring, MD 20901 (US). **MORGAN, Richard, A.** [US/US]; 6922 Tolling Bells Court, Columbia, MD 21044 (US). **FIRE, Andrew** [US/US]; c/o Carnegie Institute of Washington, 115 West University Parkway, Baltimore, MD 21210 (US). **PARRISH, Susan** [US/US]; c/o Carnegie Institute of Washington, 115 West University Parkway, Baltimore, MD 21210 (US). **MOUSSES, Spyro** [CA/US]; 257 Congressional Lane, Apt. 417, Rockville, MD 20852 (US). **KALLIONIEMI, Olli** [FI/US]; 9601 Marathon Terrace #402, Gaithersburg, MD 20878 (US). **CORNELISON, J., Robert** [US/US]; 13627 Deerwater Drive, Germantown, MD 20874 (US). **ALTON, Eric, WFW** [GB/GB]; c/o Department of Gene Therapy, Faculty of Medecine at the National Heart and Lung, Institute. Imperial College., Manresa Road, London SW3 6LR (GB). **GRIESENBACH, Uta** [DE/GB]; c/o Department of Gene Therapy, Faculty of Medecine at the National Heart and Lung, Institute. Imperial College., Manresa Road, London SW3 6LR (GB).

(74) Agent: **HARDING, Tanya, M.**; Klarquist Sparkman, LLP, One World Trade Center, Suite 1600, 121 SW Salmon Street, Portland, OR 97204 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SPECIFIC INHIBITION OF GENE EXPRESSION BY SMALL DOUBLE STRANDED RNAS

(57) Abstract: Synthetic small double stranded RNAs that can induce gene specific inhibition of expression are provided, as are methods for selecting and constructing optimal small double-stranded RNAs. Provided RNAs are suitable for interference or inhibition of expression of a target gene and comprise double stranded RNAs of about to about 40 nucleotides containing a 3' and/or 5' overhang on each strand having a length of O-nucleotide to 5-nucleotides, wherein the sequence of the double stranded RNAs are substantially identical (*e.g.*, differing by no more than 30%, or more preferably no more than 10%) to a portion of a mRNA or transcript of the target gene for which interference or inhibition of expression is desired. The current disclosure is also directed to the use of such synthetic small double stranded RNAs as reverse genetic and/or therapeutic tools. Also provided are methods, including high-throughput methods, of testing, analyzing, and optimizing short double-stranded 15 RNAs for use in gene specific inhibition of expression.

WO 03/012052 A2

SPECIFIC INHIBITION OF GENE EXPRESSION BY SMALL DOUBLE STRANDED RNAs

RELATED CASES

This application claims the benefit of co-pending U.S. Provisional Patent Applications 60/308,640, filed July 30, 2001, and 60/370,970, filed April 8, 2002, both of which are incorporated herein in their entirety.

FIELD OF THE DISCLOSURE

The current disclosure relates to synthetic small double stranded RNAs that can induce gene-specific inhibition of expression in invertebrate and vertebrate species. It is also directed to the use of such synthetic small double stranded RNAs as reverse genetic and / or therapeutic tools, and includes specific synthetic small double stranded RNAs for such uses. Many examples of specific small double stranded RNAs are also provided.

BACKGROUND OF THE DISCLOSURE

Mechanisms that silence unwanted gene expression are critical for normal cellular function. Characterized gene silencing mechanisms include a variety of transcriptional and post-transcriptional surveillance processes (Wolffe *et al.*, *Science* 286:481, 1999; Frischmeyer *et al.*, *Hum. Mol. Genet.* 8:1893, 1999; Mitchell *et al.*, *Curr. Opin. Genet. Dev.* 10:193, 2000). Double stranded RNA (dsRNA) has been shown to trigger one of these post-transcriptional surveillance processes, in which gene silencing involves the degradation of single stranded RNA (ssRNA) targets complementary to the dsRNA trigger (Fire, *Trends Genet.* 15:358, 1999). RNA interference (RNAi) effects triggered by dsRNA have been demonstrated in a number of organisms, including plants, protozoa, nematodes and insects (Cogoni *et al.*, *Curr. Opin. Genet. Dev.* 10:638, 2000). RNAi may play a role in the silencing of mobile elements in *Caenorhabditis elegans* and *Drosophila* (Kasschau *et al.*, *Cell* 95:461, 1998; Llave *et al.*, *Proc. Natl. Acad. Sci. USA* 97:13401, 2000; Tabara *et al.*, *Cell* 99:123, 1999; Ketting *et al.*, *Cell* 99:133, 1999). Similar post-transcriptional gene silencing (PTGS) effects have been implicated as an anti-viral response in plants. PTGS/RNAi appears to be a multi-step pathway requiring the processing of the trigger, a facilitated interaction with, and degradation of, the target mRNA. In some cases, these processes may also involve physical amplification of the trigger RNA and long-term maintenance of gene silencing (Matzke *et al.*, *Curr. Opin. Genet. Dev.* 11:221, 2001; Carthew, *Curr. Opin. Cell Biol.* 13:244, 2001).

A key finding from recent work has shown the generation of small (about 21 to about 25 nucleotides (nts)) dsRNAs from the input dsRNA during PTGS and RNAi (Hamilton *et al.*, *Science* 286:950, 1999; Zamore *et al.*, *Cell* 101:25, 2000; Hammond *et al.*, *Nature* 404:293, 2000; Yang *et al.*, *Curr Biol* 10:1191, 2000; Parrish *et al.*, *Mol. Cell* 6:1077, 2000). These small dsRNAs have been detected in plants, *Drosophila*, and *C. elegans* and have been suggested to serve as guide RNAs for target recognition. In *Drosophila* extracts subjected to RNAi, these short interfering RNAs (siRNAs)

resemble breakdown products of an RNase III like digestion; in particular, each strand of the siRNAs carry 5' phosphate and 3' hydroxyl termini and two or three nt 3' overhangs. SiRNAs of 21-22 nts can induce specific degradation when added to *Drosophila* cell extracts (Elbashir *et al.*, *Genes Dev.* 15:188, 2001). Further, a *Drosophila* dsRNA-specific RNase has been identified that can degrade large dsRNA (200 and 500 bp) to small dsRNAs of about 22 nts. RNAi triggered inhibition of this ribonuclease significantly reduces the effectiveness of RNAi in *Drosophila* S2 cells (Bernstein *et al.*, *Nature* 409:363, 2001).

As yet, clear evidence for the generality of an RNAi-like mechanism in vertebrate cells is lacking. Several studies have reported evidence for dsRNA-triggered silencing in particular vertebrate systems, including early embryos of mice, Zebrafish, and Xenopus, as well as Chinese hamster ovary cells (Wianny *et al.*, *Nat. Cell Biol.* 2:70, 2000; Svoboda *et al.*, *Development* 127:4147, 2000; Wargelius *et al.*, *Biochem. Biophys. Res. Commun.* 263:156, 1999; Li *et al.*, *Dev Biol* 217:394, 2000; Nakano *et al.*, *Biochem. Biophys. Res. Commun.* 274:434, 2000; Ui-Tei *et al.*, *FEBS Lett.* 479:79, 2000; Oelgeschlager *et al.*, *Nature* 405:757, 2000).

At the same time, numerous reports have described failures to observe gene-specific RNAi effects in different vertebrate systems, demonstrating instead non-specific effects of dsRNA on gene expression (Tuschl *et al.*, *Genes Dev.* 13:3191, 1999; Caplen *et al.*, *Gene* 252:95, 2000; Oates *et al.*, *Dev .Biol.* 224:20, 2000; Zhao *et al.*, *Dev. Biol.* 229:215, 2001). These non-specific effects have not been surprising, as there is an extensive literature describing a variety of non-specific responses induced by dsRNAs in mammalian cells. A major component of the mammalian non-specific response to dsRNA is mediated by the dsRNA-dependent protein kinase, PKR, which phosphorylates and inactivates the translation factor eIF2α, leading to a generalized suppression of protein synthesis and cell death via both non-apoptotic and apoptotic pathways (Clemens *et al.*, *J. Interferon Cytokine Res.* 17:503, 1997). PKR may be one of several kinases in mammalian cells that can mediate this response (Abraham *et al.*, *J. Biol. Chem.* 274:5953, 1999). A second dsRNA-response pathway involving the dsRNA-induced synthesis of 2'-5' polyadenylic acid and a consequent activation of a sequence-non-specific RNase (RNaseL) has also been demonstrated (Player *et al.*, *Pharmacol. Ther.* 78:55, 1998). The activation of PKR by dsRNA has been shown to be length dependent; dsRNAs of less than 30 nts are unable to activate PKR and full activation requires approximately 80 nts (Minks *et al.*, *J. Biol. Chem.* 254:10180, 1979; Manche *et al.*, *Mol. Cell Biol.* 12:5238, 1992).

SUMMARY OF THE DISCLOSURE

We have now determined that small (about 15 to about 40 nucleotides (nts), in some embodiments about 20 to about 25 nts) dsRNAs can trigger a gene-specific interference or inhibition of gene expression in model invertebrates and mammalian cells without activating PKR. Such synthetic small double stranded RNAs can be used as reverse genetic and/or therapeutic tools.

The current disclosure relates to synthetic small double stranded RNAs that can induce gene-specific interference or inhibition of expression in invertebrate and vertebrate species. It is also

directed to the use of such synthetic small double stranded RNAs as reverse genetic and/or therapeutic tools.

Small double stranded RNAs (dsRNAs) of this disclosure are double stranded RNAs of about 15 to about 40 nts, in some embodiments of about 20 to about 25 nts, that interfere with, or inhibit, expression of a target sequence. These double stranded RNAs have 0-nucleotide (*i.e.*, blunt ends) to 5-nucleotide 3' and/or 5' overhangs on each strand of the duplex; the RNA backbone and/or component nucleosides may be unmodified or modified. The sequence of the small double stranded RNA is substantially identical to a portion of a mRNA or transcript of a target gene for which interference or inhibition of expression is desired. Synthetic small dsRNAs can induce gene-specific inhibition of expression in *Caenorhabditis elegans, Drosophila melanogaster* embryonic cells, and in cell lines from humans and mice, for instance. In each case, the interference by small dsRNAs was superior to the inhibition of gene expression mediated by single stranded antisense oligoribonucleotides. Provided small dsRNAs appear to avoid the well-documented, non-specific effects triggered by longer dsRNAs in mammalian cells. Thus, such small dsRNAs may be used as reverse genetic and therapeutic tools in mammalian cells, including human cells, both *in vitro* and *in vivo*.

The present disclosure in several embodiments provides RNA suitable for interference or inhibition of expression of a target, which RNA includes double stranded RNA of about 15 to about 40 nts containing a 0-nucleotide to 5-nucleotide 3' and/or 5' overhang on each strand, wherein the sequence of the double stranded RNA is substantially identical to a portion of a mRNA or transcript of the target for which interference or inhibition of expression is desired. The RNA backbone and/or component nucleosides of the small dsRNAs of this invention may be unmodified or modified.

Another embodiment is a method for treating animal cells by interfering with or inhibiting expression of a target in the animal cells, which method involves exposing the animal cells to an effective amount of RNA suitable for interfering or inhibiting expression of the target, wherein the RNA includes double stranded RNA of about 15 to about 40 nucleotides containing a 0-nucleotide to 5-nucleotide 3' and/or 5' overhang on each strand, wherein the sequence of the double stranded RNA is substantially identical to a portion of the transcript (mRNA) encoded by the target gene. The RNA backbone and/or component nucleosides of the small dsRNAs of this invention may be unmodified or modified. Such methods are especially suitable for the treatment of mammal cells, more especially for the treatment of human cells *in vivo* and *in vitro*.

The present disclosure also provides a genetic method for treating an animal by interfering with or inhibiting expression of a target sequence in the animal, said method comprising administering to the animal an effective amount of RNA suitable for interfering or inhibiting expression of the target sequence, wherein the RNA comprises double stranded RNA of about 15 to about 40 nucleotides containing a 0-nucleotide to 5-nucleotide 3' and/or 5' overhang on each strand, wherein the sequence of the double stranded RNA is substantially identical to a portion of a mRNA or transcript of the target sequence. The RNA backbone and/or component nucleosides of the small

dsRNAs of this invention may be unmodified or modified. This method is especially suitable for the treatment of mammals, more especially for the treatment of humans.

Also provided herein are methods for optimizing small double-stranded RNAs for interference with or inhibition of expression of a target gene, as well as collections of optimized molecules, including high-throughput and array-based methods. Small double-stranded RNAs are provided that can be used to inhibit or interfere with the expression of specific genes, including *Erb-B2*, *APPBP2*, *BMP7*, *CCND1*, *CRYM*, *ER1*, *FKBP5*, *FLJ20940*, *GRB7*, *HOXB7*, *LMO4*, *MGC9753*, *MLN64*, *MYBL2*, *MYC*, *NBS1*, *NCOA3*, *PIP5K2B*, *PNMT*, *PPARBP*, *PPM1D*, *RAD51C*, *RAE1*, *RPS6K*. *S100P*, *TBX2*, *TMEPAI*, *TRIM37*, *TXNIP*, and *ZNF217*, each of which is associated with cancer.

The foregoing and other features and advantages will become more apparent from the following detailed description of several embodiments, which proceeds with reference to the accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a pair of bar graphs, illustrating effects of small double stranded RNAs (SEQ ID NOs: 1-14, 39, 40, 45 and 46) on gene expression in *Drosophila* embryonic S2 cells. *Drosophila* S2 cells co-transfected with pAct.GFP and control and *egfp* dsRNAs were assessed 48 hours after transfection for GFP expression by fluorescence activated cell sort (FACs) analysis. Fig. 1A shows the percentage of GFP positive cells; Fig. 1B shows the fluorescence intensity (geometric mean) of GFP positive cells. Each transfection was assayed in triplicate and data is shown as a mean ± SEM. * indicates $P<0.05$; ** indicates $P<0.01$; and *** indicates $P<0.001$.

Figure 2 is a series of graphs and gels, showing gene-specific inhibition of expression in MEFs by small double stranded RNAs. MEFs transfected with plasmid DNA and ssRNAs and dsRNAs (SEQ ID NOs: 1-25, 39,40, 45 and 46) were harvested 48 hours after transfection and assayed for:

Fig. 2A - 2D -- GFP expression by FACs analysis (each transfection was assayed in triplicate and data is shown as mean ± SEM). Fig. 2A and 2C show the percentage of GFP positive cells, Fig. 2B and 2D show the fluorescence intensity (geometric mean) of GFP positive cells.

Fig. 2E - 2I -- CAT expression (each transfection condition was assayed in triplicate). Fig. 2E data is normalized to the amount of CAT pg/μg protein observed in plasmid only, (F-I) data is normalized to the amount of CAT pg/μg protein in plasmid and sense ssRNA transfected cells, (s = sense ssRNA; as = antisense ssRNA; ds = dsRNA).

Fig. 2J -- *egfp* and *neo* RNA levels by Northern analysis of poly A+ mRNA.

Fig. 2K -- Cell survival (assayed in duplicate and shown as a mean $OD_{560/650}$), dsRNAs of 21-25 and 78 nts correspond to *egfp*, the dsRNA of 81 nts corresponds to *LacZ*.

Fig. 2L and 2M -- GFP expression by FACs analysis (data is shown as relative percentage normalized to pEGFP-N3 transfected cells).

* Indicates $P<0.05$; ** indicates $P<0.01$; and *** indicates $P<0.001$.

Figure 3 is a series of bar graphs showing small double stranded RNA (SEQ ID NOs: 1-6, 15 – 18) mediated gene silencing in human cells.

293 (Fig. 3A and 3B) and HeLa (Fig. 3C and D) cells transfected with pEGFP-N3 and antisense (as) ssRNAs and dsRNAs were harvested 48 hours after transfection and assayed for GFP expression by FACs analysis (assayed in triplicate, data is shown as mean ± SEM). Fig. 3A and 3C show the percentage of GFP positive cells, Fig. 3B and 3D show the fluorescence intensity (Geo Mean) of GFP positive cells.

For Fig. 3E - 3H, HeLa cells transfected with pcDNA3-CAT and ssRNAs and dsRNAs were harvested 48 hours after transfection and assayed for CAT expression (assayed in triplicate and normalized to the amount of CAT pg/μg protein observed in plasmid plus sense transfected cells). s = sense ssRNA; as = antisense ssRNA; ds = dsRNA, * indicates $P<0.05$; ** indicates $P<0.01$; and *** indicates $P<0.001$.

Figure 4 is a series of gels and bar graphs, illustrating that mammalian cells have small dsRNA and dsRNA-dependent pathways. *In vitro* kinase assays were performed to detect PKR autophosphorylation.

For Fig. 4A, *in vitro* kinase reactions were performed without exogenous RNA (-), or with 1 μg/ml reovirus dsRNA or 1 μg/ml small dsRNA (21-25 nts; SEQ ID NOs: 1-10), or 1 μg/ml 78 or 81 nt dsRNA (SEQ ID NOs: 39, 40, 45-46). For Fig. 4B, *in vitro* kinase competition assays were performed using small and large dsRNAs. Reactions were performed without exogenous RNA (-), or 1 μg/ml reovirus RNA, or 75 fold excess small dsRNA (21-25 nt), or 78 or 81 nt dsRNA, plus reovirus dsRNA (1 μg/ml). Small dsRNAs of 21-25 nts and dsRNA of 78 nts corresponded to *egfp*, the 81 nt dsRNA corresponds to *LacZ*).

293 cells transfected with pEGFP-N3 and *unc-22* (SEQ ID NOs: 19-22) or *egfp* small dsRNAs (SEQ ID NOs: 5-8) (Fig. 4C and 4D) and 293 cells transfected with pEGFP-N3 and 78 *egfp* dsRNA and *unc-22* or *egfp* small dsRNAs (Fig. 4E and 4F) were assayed for GFP expression by FACs analysis 48 hours after transfection (each transfection was assayed in triplicate, data is shown as mean ± SEM). Fig. 4B and 4D show the percentage of GFP positive cells; Fig. 4C and 4E show the fluorescence intensity (Geo Mean) of GFP positive cells). * indicates $P<0.05$; ** indicates $P<0.01$; and *** indicates $P<0.001$.

Figure 5 is a series of panels, illustrating inhibition of stable d2eGFP expression in human HeLa (HeLa/d2eGFP) cells following transfection with small dsRNAs.

Fig. 5A is a diagrammatic representation of the relative position of four dsRNAs (21-24 nucleotides in length) corresponding to different regions of the *egfp* encoding sequence (Region A, SEQ ID NOs: 1 and 2; Region B, SEQ ID NOs: 3 and 4; Region C, SEQ ID NOs: 5 and 6; Region D, SEQ ID NOs: 7 and 8).

Fig. 5B is a series of six representative FACs histograms of HeLa/d2eGFP cells analyzed 48 hours post transfection with dsRNAs corresponding to four different regions of the *egfp* transcript and varying in size from 21 to 24 nts (i-vi) or dsRNAs corresponding to a region of a control gene, ß-galactosidase (*Lac-Z*; SEQ ID NOs: 29-32) of 22 or 23 nts in length (v and vi). In each histogram,

the light gray line shows the FACs histogram for control HeLa cells, the medium gray line a typical FACs histogram analysis for the HeLa/d2eGFP cell population, and the solid gray histogram eGFP expression in cells transfected with the stated dsRNA.

Fig. 5C is a pair of bar graphs showing quantification of FACs analysis (percentage GFP positive cells and the fluorescence intensity of all cells assessed by the geometric mean of all cells) of HeLa/d2eGFP cells transfected with the egfp dsRNAs shown in (Fig. 5A) or two control dsRNAs against ß-galactosidase, the data is from 3 independent transfections for each dsRNA and is shown as mean ± SEM.

Figure 6 is a series of panels, illustrating that both the size and the sequence of an siRNA are involved in determining the effectiveness of an siRNA against egfp.

Fig. 6A is a series of seven representative FACs histograms (light gray line = FACs histogram for control HeLa cells, the medium gray line = FACs histogram for the HeLa/d2eGFP cell population, the solid gray histogram = eGFP expression in cells transfected with the stated dsRNA).

Fig. 6B is a pair of bar graphs, showing quantitative data from Hela/d2eGFP cells transfected with dsRNAs (3 independent transfections for each dsRNA, mean ± SEM) corresponding to one region of the egfp transcript but varying in size from 20 to 24 nts. The illustrated oligonucleotides have sequences as follows: *egfp* region D 20, SEQ ID NOs: 37 and 38 and residues 1-20 of SEQ ID NOs: 7 and 8; *egfp* region D 21, residues 1-21 of SEQ ID NO: 7 and residues 1-21 of SEQ ID NO: 8; *egfp* region D 22, residues 1-22 of SEQ ID NO: 7 and residues 1-22 of SEQ ID NO: 8; *egfp* region D 23, residues 1-23 of SEQ ID NO: 7 and residues 1-23 of SEQ ID NO: 8, and *egfp* region D 24, SEQ ID NO: 7 and SEQ ID NO: 8.

Fig. 6C shows alignment of *egfp* dsRNAs spanning region B (which are listed in Table 3, SEQ ID NOs: 107-110, 99, 100, 97, 98, 4, 112, and 115-132) on two different m-folds of the egfp transcript. The *egfp* small dsRNAs shown cover a 50 nucleotide region of the *egfp* transcript that includes the "Region B" sequence (SEQ ID NOs: 3 and 4); in this figure, the symbol "#" correspond to a region where the majority of small dsRNAs effectively inhibit eGFP expression, "+" correspond to a region where the small dsRNAs against this region mediate an intermediate level of inhibition of gene expression, and "*" correspond to a region where the small dsRNAs fail to mediate inhibition of gene expression.

Figure 7 is a bar graph and a series of representative images showing that dsRNAs corresponding to the transgene ß-galactosidase (LacZ Sense Z1 – Z10; SEQ ID NOs: 148, 150, 152, 154, 156, 158, 160, 162, 164, and 166) can interfere with the expression of this gene in mouse 3T3 fibroblasts.

The mouse fibroblast cell line 3T3 cells permanently expressing the *E. coli* ß–galactosidase (*LacZ*) gene were transfected with small dsRNAs corresponding to different regions of this gene (see Table 3 for sequences; LacZ Sense Z1 – Z10; SEQ ID NOs: 148, 150, 152, 154, 156, 158, 160, 162, 164, and 166) and control small dsRNAs corresponding to *cat* or *egfp* (see Table 3 for sequences, SEQ ID NOs: 15 and 16, 3 and 4). Cells were analyzed 72 hours after transfection using either (Fig.

7A) a commercially available luminescent ß-galactosidase assay (Clontech Inc.) or, (Fig. 7B) by microscope analysis using standard X-gal staining.

Figure 8 is a pair of bar graphs, showing that small double stranded RNAs with different termini can interfere with or inhibit gene expression in mouse embryonic fibroblasts. Mouse embryonic fibroblasts were co-transfected with pEGFP-N3 and small dsRNAs corresponding to different portions of the *egfp* sequence and with different termini. Cells were harvested 48 hours after transfection and assayed for GFP expression by FACs analysis (assayed in triplicate, data is shown as mean ± SEM). The percentage of GFP positive cells is shown in Fig. 8A. The fluorescence intensity (Geometric Mean) of GFP positive cells is shown in Fig. 8B. The small dsRNAs used corresponded to 'Region B' of the *egfp* mRNA and have different 5' and 3' overhangs (see Table 3 SEQ ID NOs: 3 and 33, 3 and 34, 3 and 35 and 3 and 36) and

Figure 9 is a bar graph illustrating that small double stranded RNAs with different termini can interfere with, or inhibit, pre-existing stable gene expression in human cells (quantified as percentage GFP positive cells). Human 293 cells were transfected with a mammalian expression plasmid that encodes a modified version of EGFP that has a half-life of approximately 2 hours (pEGFPd2) and the neomycin phosphotransferase genes. These cells were selected for resistance to the neomycin phosphotransferase analog G418 (1 mg/ml) for several weeks to ensure the stable expression of the EGFP protein. These cells were transfected twice with 3 μg of a given small dsRNA, the two transfections were conducted 48 hours apart. Cells were assayed for GFP expression by FACs analysis 24 hours after the second transfection. The small dsRNAs used corresponded to 'Region B' of the *egfp* mRNA and have different 5' and 3' overhangs (see Table 3 SEQ ID NOs: 3 and 33, 3 and 34, 3 and 35 and 3 and 36) and

Figure 10 illustrates effects of the size and nucleotide composition of the 3' overhang on the ability of a 22 nt small dsRNA to mediate gene silencing.

Fig. 10A is a series of representative FACs histogram (i-v) from HeLa/d2eGFP cells transfected with dsRNAs corresponding to 'Region B' of the *egfp* mRNA with different 5' and 3' overhangs (see Table 3 SEQ ID NOs: 3 and 33, 3 and 34, 3 and 35 and 3 and 36) and bar graphs showing the percentage of GFP positive HeLa/d2eGFP cells (vi) and the fluorescence intensity (geometric mean) (vii)

Fig. 10B is a series of eight bar graphs, showing the percentage of GFP positive HeLa/d2eGFP cells (i, iii, v, vii) and the fluorescence intensity (geometric mean, ii, iv, vi, viii) of HeLa/d2eGFP cells transfected with small dsRNAs formed from four different antisense ssRNAs (i and ii, antisense strand SEQ ID NO: 112, iii and iv, antisense strand SEQ ID NO: 114, v and vi, antisense strand SEQ ID NO: 118, vii and viii, antisense strand SEQ ID NO: 124; also see Table 3 for sequences) combined with different sense single stranded oligoribonucleotides (SEQ ID NOs: 3, 107, 97, 111, 113, 117, 119, 121, 125, 127, and 131 – also see Table 3 for sequences) to generate dsRNAs with overhangs ranging from 6 nts 5' to 8 nts 3'.

Fig. 10C and 10D are bar graphs showing the percentage of GFP positive Hela/d2eGFP cells (i) and the fluorescence intensity (geometric mean, ii) of Hela/d2eGFP cells transfected with small

dsRNAs against the Region B of the egfp mRNA where the original 2 nt 3' overhang nucleotides were replaced by (Fig. 10C) a deoxythymidine doublet (dTdT) or (Fig. 10D) a uridine doublet (UU) (SEQ ID NOs: 133-136).

Figure 11 relates to effects of backbone modifications on the activity of an RNAi dsRNA in *C. elegans*. Fig. 11A is schematic drawing of an RNA backbone, including a description of the modifications used in this work. The uracil-to-deoxythymidine substitution also entails a change in the base moiety (see Figure 11A).

Fig. 11B is a table, summarizing activities of backbone-modified RNAs. The degree of interference was assessed by examination of phenotypes of progeny animals with and without levamisole treatment. Although assays were in general qualitative, we have included an indicator of interference strength based on severity of phenotypes for the majority class of progeny animals scored. Unmodified RNA and modified RNAs described as "+++" produced a strong twitching phenotype in the absence of levamisole (Fire *et al., Nature* 391 806, 1998). Modified RNAs described as "++" produce a weak twitching phenotype that is evident without levamisole treatment as a twitch during movement; these animals twitch strongly following levamisole treatment. Interference described as "+" was observed only in levamisole and only in a fraction of progeny. Interference described as "+/-" was at a level that was indistinguishable from the preparations of unmodified sense or antisense RNA used to make the modified duplex. Titrations of unmodified RNA (Fire *et al., Nature* 391 806, 1998) indicate that levels of twitching described as ++ correspond to activities in a range of 3–10 percent of unmodified RNA, while those described as "+" correspond to 1–3 percent of unmodified RNA. 2'-fluorouracil, 2'-aminouracil, 2'-deoxythymidine, and 2'-deoxycytidine were incorporated into individual strands of the 742 nt *unc-22*A segment using T3 and T7 RNA polymerases. Incorporation of α-thio nucleotides was carried out with an RNA hairpin carrying *unc-22* sequences in the stem. This allows rapid formation in the polymerize reaction mix of duplex structure, which appeared to stabilize the resulting modified RNA. RNA concentrations in injection mixes were in the range of 30–60 mg/ml.

Figure 12 relates to effects of base modifications on the activity of an RNAi dsRNA in *C. elegans*. Fig. 12A shows the formulas for uracil and guanosine, with a description of the modifications used in this work.

Fig. 12B is a table summarizing activities of backbone modified RNAs.

Fig. 12C is a series of three bar graphs and a table, showing activities of 5-(3-aminoallyl)-uracil substituted RNAs. The triggering segments *unc22A*, *gfpG*, *lacZL*, and *unc54A* were as described in Fire *et al.*, (*Nature* 391 806, 1998. RNA concentrations were 30 mg/ml for *lac*ZL and *gfp*G, and 40 mg/ml for *unc-22*A. *gfp*G and *lac*ZL injections were carried out in strain PD4251 (Fire *et al., Nature* 391 806, 1998). *unc-22*A and *unc-54A* injections were carried out in wild type (N2) animals. For *unc-22*A injections, twitching fractions shown are in the absence of levamisole (L4 larvae and young adults). Some animals with weak twitching in the presence of levamisole were observed in antisense-modified *unc-22*A injections (57/200 animals); this was only marginally above the background level of weak levamisole-induced twitching following injection of the preparation of

sense RNA alone (30/200 animals); this limited signal could represent residual contamination of dsRNA in the individual ssRNA preparations. For *unc-54*A injections, movement was scored in L1 larvae and again in young adults. Titrations of 5-(3-aminoallyl)uracil sense-modified RNAs for *unc-54*A indicate that this RNA is at least 15- fold more active than the equivalent antisense modified *unc-54*A trigger. Similar experiments with *lacZ*L, *gfp*G, and *unc-22*A indicate in each case at least a 10-fold differential strand effect for 5-(3-aminoallyl)uracil substitution.

Figure 13 demonstrates that there is an asymmetric role for the sense and antisense strands in RNAi in mammalian cells, based on chemical modification of the component single stranded oligoribonucleotides.

Fig. 13A is a series of bar graphs (i and ii) and representative FACs histogram (iii-vi) data from HeLa/d2eGFP cells transfected with dsRNAs corresponding to 'Region B' of the *egfp* mRNA with 2'O methyl modifications of the RNA backbone (SEQ ID NOs: 137 and 138); (i and ii) 3 independent transfections for each dsRNA, mean ± SEM) and representative FACs histograms (iii – vi, light gray line = FACs histogram for control HeLa cells, medium gray line = FACs histogram for the HeLa/d2eGFP cell population, solid dark gray histogram = eGFP expression in cells transfected with the stated dsRNA).

Fig. 13B is a series of bar graphs (i and ii) and representative FACs histogram (iii-vi) data from HeLa/d2eGFP cells transfected with dsRNAs (SEQ ID NOs: 3 and 4) corresponding to 'Region B' of the *egfp* mRNA with or without a 5' phosphate group (i and ii) 3 independent transfections for each dsRNA, mean ± SEM) and representative FACs histograms (iii – vi, light gray line = FACs histogram for control HeLa cells, medium gray line = FACs histogram for the HeLa/d2eGFP cell population, solid dark gray histogram = eGFP expression in cells transfected with the stated dsRNA).

Fig. 13C is a series of bar graphs (i and ii) and representative FACs histogram (iii-vi) data from HeLa/d2eGFP cells transfected with dsRNAs corresponding to 'Region B' of the *egfp* mRNA with or without a 5' Rhodamine (TMARA) fluorophore group (i and ii) 3 independent transfections for each dsRNA, mean ± SEM) and representative FACs histograms (iii – vi, light gray line = FACs histogram for control HeLa cells, medium gray line = FACs histogram for the.HeLa/d2eGFP cell population, solid dark gray histogram = eGFP expression in cells transfected with the stated dsRNA).

Fig. 13D is a series of bar graphs (i and ii) and representative FACs histogram (iii-vi) data from HeLa/d2eGFP cells transfected with dsRNAs corresponding to 'Region B' of the *egfp* mRNA with or without a 3' Rhodamine (TMARA) fluorophore group (i and ii) 3 independent transfections for each dsRNA, mean ± SEM) and representative FACs histograms (iii – vi, light gray line = FACs histogram for control HeLa cells, medium gray line = FACs histogram for the HeLa/d2eGFP cell population, solid dark gray histogram = eGFP expression in cells transfected with the stated dsRNA) (SEQ ID NOs: 3, 4, 143, and 144).

Figure 14 is a line graph, showing that RNAi in HeLa/d2eGFP cells is transient, but can be re-established. HeLa/d2eGFP cells were transfected with either a control small dsRNA (*LacZ;* SEQ ID NOs: 29 and 30) or an *egfp* (SEQ ID NOs: 3 and 4) small dsRNA. Cells were subjected to FACs

analysis at the time points shown. after 12 days calls were retreated with the same dsRNAs. The data shown is the mean ± SEM for each treatment at each time point.

Figure 15 is a series of panels, showing that variations in the sequence of the small dsRNA (SEQ ID NOs: 3 and 145) can be used to target all of the alternative sequences forms of the particular transcript, or they may be a used to specifically distinguish between these transcripts.

Fig. 15 shows a series of bar graphs (i and ii; 3 independent transfections for each dsRNA, mean ± SEM) and representative FACs histogram (iii-vi, light gray line = FACs histogram for the HeLa/d2eGFP cell population transfected with a control (cat) dsRNA, solid gray histogram = eGFP expression in cells transfected with the stated dsRNA). See table 3 and Fig. 15 for sequences. As with other modifications of the component strands of the dsRNA we saw an asymmetric effect of the three-nucleotide change. When the nucleotide change was present on the sense strand there was no effect on the ability of the dsRNA to mediate RNAi, however, when on the antisense strand the percentage of cells inhibited was significantly less. This suggests that small (between 1-10 nt) variations in sequence could be used to distinguish between transcripts or to ensure that all of the sequence variants of a particular transcript are targeted. The three-nucleotide change used here was incorporated in approximately the middle of the dsRNA sequence, but variations in the sequence could be incorporated into any position within the dsRNA.

Figure 16 is a pictoral overview of RNAi microarray technology, showing the inter-relation of small dsRNA design, microarray manufacturing, live-cell array analysis, multiple end-point analyses, data extraction, and a web-based information management system. This illustrates one embodiment and it is understood that other systems could be used, wherein the details could vary from those illustrated.

Figure 17 is a series of bar graphs, illustrating an assessment of the sensitivity and specificity of the RNAi microarray platform.

Fig. 17A is an analysis of eGFP expression (median green fluorescence intensity) in HeLa/d2eGFP cells over time following exposure to either arrayed *egfp* (black bars; SEQ ID NOs: 3 and 4) or negative control (*cat*; SEQ ID NOs: 15 and 16) dsRNAs (grey bars) (2.5 ng dsRNA per spot). The inhibition of eGFP expression by arrayed *egfp* dsRNAs (black bars) is dose dependent (0.25 ng, 1 ng or 2.5 ng dsRNA) and this effect can be observed over time ((Fig. 17B) 24 hours (Fig. 17C) 72 hours). No change was seen in the fluorescence intensity of Hela/d2eGFP cells exposed to the negative control, *cat*, dsRNA (grey bars).

Fig. 17D shows that cells exposed to arrayed dsRNAs (negative control, grey bars; experimental dsRNA, black bars) were of the same size and were otherwise comparable; see Addenda A and B and Figure 12 for additional statistics. Three different dsRNAs against *egfp* were compared for their ability to inhibit eGFP expression RNAi microarray (2.5 ng dsRNA) (Fig. 17E) and flow cytometry (2 μg dsRNA was transfected using Lipofectin into approximately $2x10^5$ HeLa/d2eGFP cells) (Fig. 17F).

In Fig. 17A- Fig. 17E, the average median green fluorescence of triplicate dsRNA spots is plotted with the error bar representing one standard deviation, in (Fig. 17F) the average geometric

mean green fluorescence of three independent transfections is plotted with the error bar representing one standard deviation.

Figure 18 is a series of panels, illustrating *Drosophila* plasmids expressing AR.CAG$_n$AR.GFP fusion transcripts and their transient expression in S2 cells.

Fig. 18A is a diagrammatic representations of *Drosophila* expression plasmids used in this study.

Fig. 18B is a micrograph, illustrating fluorescence in S2 cells (2 x10^6) that were transfected with pAct.GFP, pAct.ARCAG26GFP, pAct.ARCAG43GFP, and pAct.ARCAG106GFP (2 μg plasmid DNA) and were analyzed by deconvolution microscopy 72 hours after initiation of transfection, GFP fluorescence is in green and nuclei are countered stained in blue using Hoeschst staining.

Figure 18C is a diagrammatic representation of the dsRNAs used in this study related to the arcag$_n$ar(±gfp) fusion transcript.

Figure 19 illustrates inhibition of transgene expression by gene-specific dsRNA molecules. Double stranded RNA molecules corresponding to *LacZ*, *cat*, *gfp*, or the androgen receptor, *ar* were co-transfected into S2 cells with either (Fig. 19A) pAct.ARCAG26GFP, (Fig. 19B) pAct.ARCAG43GFP (Fig. 19C) pAct.ARCAG106GFP, (Fig. 19D) pAct.GFP, or (Fig. 19E) pAct.CAT. S2 cells (2 x10^6) were transfected with 2 μg of plasmid DNA and 1 μg of dsRNA; cells were analyzed 72 hours after initiation of transfection. Data shown are representative of at least three independent experiments using each dsRNA and is normalized to the levels of GFP expression observed in those cells co-transfected with the plasmid shown and the control *LacZ* dsRNA, n=3 for each plasmid/dsRNA combination.

Fig. 19F is a Northern analysis of Poly A+ RNA (approximately 2 μg of each sample) purified from S2 cells transfected with pAct.GFP, pAct.ARCAG26GFP, pAct.ARCAG43GFP and pAct.ARCAG106GFP and either *LacZ*, *cat*, *gfp* or *ar* dsRNA. The Northern blot was sequentially hybridized with (i) human *ar*, (ii) *gfp*, and (iii) *Drosophila gapdh-1* specific DNA probes. Statistical analysis was performed by comparison to *LacZ* dsRNA transfected cells ** $p<0.01$, *** <0.001.

Figure 20 illustrates the effect of CAG containing dsRNA molecules on arcag$_n$ar fusion transcripts.

S2 cells were transfected with pAct.ARCAG26GFP (Fig. 20A & Fig. 20E), pAct.ARCAG43GFP (Fig. 20B & Fig. 20F), pAct.ARCAG106GFP (Fig. 20D & Fig. 20H) or pAct.GFP (Fig. 20D & Fig. 20H) and dsRNA molecules corresponding to: *LacZ* (Fig. 20A-20H), 27 CAG trinucleotide repeats (*cag27*), (Fig. 20A-20D), 21nts from the upstream flanking sequence of the human androgen receptor plus 27 CAG repeats (*arcag27*) (Fig. 20-20D), *arcag26ar* (Fig. 20E-20H) *arcag43ar* (Fig. 20E-20H), *arcag106ar* (Fig. 20E-20H) or *gfp* (Fig. 20E-20H). S2 cells (2 x10^6) were transfected with 2 μg of plasmid DNA and 0.5-1 μg of dsRNA; cells were analyzed 72 hours after initiation of transfection, (Fig. 20A-20D) n=3, E-H (n=5), data shown are normalized to the levels of GFP expression observed in those cells co-transfected with the plasmid shown and the

control *LacZ* dsRNA. Statistical analysis was performed by comparison to *LacZ* dsRNA transfected cells ** $p<0.01$, *** <0.001.

Figure 21 illustrates the specific inhibition of $arcag_nar$ transcripts by small dsRNAs in human 293 cells and rescue of CAG induced cytotoxicity.

Fig. 21A-20H, 293 cells (1 x 10^6) were transfected with pCMV.ARCAG112GFP (Fig. 21A, 20C, 20E & 20G) or pCMV.ARCAG19GFP (Fig. 21B, 21D, 21F, & 21H) and dsRNAs corresponding to *cat* (Fig. 21C, 21D, 20G & 21H) or *gfp* (Fig. 21E-21H) and were assayed for GFP expression and general cell morphology by deconvolution microscopy (Fig. 21A-21F) (GFP fluorescence is in green and nuclei are countered stained in blue using Hoeschst staining) or cell death by propidium iodide (PI) staining (Fig. 21G–21H; data is shown as mean±SD, n= 3) 72 hours after initiation of transfection.

For Fig. 21I-21L, 293 cells (1 x 10^6) were transfected with pCMV.ARCAG112 (Fig. 21I & 21K) or pCMV.ARCAG16 (Fig. 21J & 21L) and dsRNAs corresponding to *cat* and two different regions of the human androgen receptor (*5'ar* and *3'ar*); cells were assayed 72 hours after initiation of transfection for cell death by propidium iodide (PI) staining (Fig 21I & 21J) and caspase-3 activity (relative fluorescence units of AFC released/ minute/μg protein) (21K & L). Data shown are mean ±SD from 3 independent transfections.

Figure 22 illustrates the effect of a small dsRNA against *erb-b2* on Erb-B2 protein levels in a SKBR3 cell line. SKBR3 is a breast cancer cell line containing a large chromosomal amplification that includes the *Erb-B2* gene and significantly over-expresses the *Erb-B2* transcript. Here we demonstrate the ability of a small dsRNA to down-regulate the expression of Erb-B2 at the protein level using (A) Western blot analysis and (B) FACs analysis. A control small dsRNA against the *LacZ* gene had no effect on Erb-B2 expression, whereas the small dsRNA targeting the *Erb-B2* mRNA significantly down-regulated Erb-B2 protein expression. The levels of a control protein (Tubulin) were unaffected by any of the small dsRNAs, however, a significant decrease in the levels of a protein (Cyclin-D) was seen which is consistent with the dependence of the expression of this protein on the presence of Erb-B2.

Figure 23 shows fluorescent microscopy images of the small dsRNA down-regulation of (A) Erb-B2 expression in SKBR3 assayed by immunofluorescence at 10x magnification with a nuclear counterstain (DAPI) by an *erb-b2* specific small dsRNA and (B) a significant reduction in the expression of the Ki67 antigen which was used to assay rates of cellular proliferation following transfection with an *erb-b2* specific small dsRNA but not a control small dsRNA.

Figure 24 illustrates Erb-B2 expression in SKBR3 cells transfected with various small dsRNAs, subjected to immunohistochemistry for Erb-B2 expression. The levels of Erb-B2 expression were quantitated from robotically captured immunofluorescent images by a custom segmentation algorithm. Images from the Erb-B2 immunofluorescence assay were acquired and run through a segmentation and quantitation routine that analyzes pixel intensities on a cell by cell basis, and generates multiparametric statistical data. The graph represents the percentage of Erb-B2 positive

cells at any mean threshold intensity. In this graph the data for four different *Erb-B2* small dsRNAs (ERB9, SEQ ID NOs: 63 and 64, Erb10, SEQ ID NOs: 65 and 66, ERB 11, SEQ ID NOs: 67 and 68, ERB14, SEQ ID NOs: 73 and 74) is shown all of which show a significant decrease in Erb-B2 expression relative to the negative control (untransfected or *egfp* small dsRNA transfected cells SEQ ID NOs: 3 and 4).

Figure 25 illustrates that a small dsRNA against *Erb-b2* induce apoptosis in the SKBR3 breast cancer cell line. The effect of dsRNA against Erb-b2 on rates of apoptosis was assayed in breast cancer cells using the Tunel DNA fragmentation assay (Clontech Inc, Apoalert). A significant increase in the number of apoptotic cells was seen following treatment with two different Erb-B2 small dsRNAs (*ERB0* SEQ ID NOs: 178 and 179 and *ERB5* SEQ ID NOs: 55 and 56)

Figure 26 shows the quantification of the rates of cellular proliferation measured by immunofluorescence using an antibody against Ki-67 in SKBR3 cells. The images were quantitated using the algorithm described in Fig. 25 and shows that multiple small dsRNAs against different target genes can be assayed at the same time. Illustrative data is shown using small dsRNA's against RPS6K, TBX2, and GRB7 (see Table 8), which all decreased the rate of proliferation compared to a negative control.

Figure 27 shows quantification of the apoptotic rate of prostate cancer cells (pc3) challenged with increasing concentrations of a small dsRNA against the anti-apoptotisis gene BCL-Xl1 (see Table 8). Quantification was performed using the method described in Fig. 25. The apoptotic rate of the PC3 cell line was determined by multiple negative controls and a dose response was apparent when treated with a small dsRNA against BCL-XL in increasing amounts. The image inset shows the 5 μg dose of dsRNA vs. the negative control. The top two images are the nuclear counterstains showing the overall confluency, and the bottom two images are the Tunel apoptosis assay where apoptotic nuclei fluorescent in an intense, puctant dark stain.

Figure 28 is a bar graph showing gene silencing through RNA interference in the murine lung *in vivo*.

The expression plasmid pCI-CAT carrying a chloramphenicol acetyl transferase reporter (*cat*) gene or an empty plasmid control (pCI) were complexed with cationic lipids (80 mg/mouse in a total volume of 100 ml). Double strand small dsRNAs corresponding to *cat* (SEQ ID NOs: 15 and 16) or *egfp* (SEQ ID NOs: 3 and 4) were complexed plasmid DNA and dsRNA were mixed prior to *in vivo* transfection. For lung transfection Balb/C mice (female, 6-8 weeks) were anaesthetized with an inhaled anesthetic (metofane) and the liposome complexes were applied to the nose and "sniffed" voluntarily into the lung. Forty-eight hours after transfection the lungs were harvested, and homogenized. CAT expression was measured with a CAT ELISA according to manufacturers recommendations. (n=10-12,**=p<0.001).

SEQUENCE LISTING

The nucleic and amino acid sequences listed in the accompanying sequence listing are shown using standard letter abbreviations for nucleotide bases, and three letter code for amino acids,

as defined in 37 C.F.R. 1.822. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand. Some of the sequences depicted herein are shown in 5' to 3' orientation; however, some are shown in the 3' to 5' orientation, for instance to emphasize annealing positions between pairs of oligoribonucleotides as shown in Table 4. In particular, certain of the antisense oligonucleotide are depicted 3' to 5'; these are indicated by adding the letter "R" (reverse) to the SEQ ID NO. In the accompanying sequence listing:

SEQ ID NOs: 1-14 and 33-38 are *egfp* sense and antisense oligonucleotides for generating small dsRNA, as detailed more fully in Table 1.

SEQ ID NOs: 15-18 are *cat* sense and antisense oligonucleotides for generating small dsRNA, as detailed more fully in Table 1.

SEQ ID NOs: 19-24 are *unc22* sense and antisense oligonucleotides for generating small dsRNA, as detailed more fully in Table 1.

SEQ ID NOs: 25-28 are *Zr unc22* sense oligonucleotides for generating small dsRNA, as detailed more fully in Table 1.

SEQ ID NOs: 29-32 are *LacZ* sense and antisense oligonucleotides for generating small dsRNA, as detailed more fully in Table 1.

SEQ ID NOs: 39-46 are *egfp* (39, 40), *cat* (41, 42), *unc22* (43, 44), and *LacZ* (45, 46) sense and antisense oligonucleotides for generating longer dsRNA, as detailed more fully in Table 1.

SEQ ID NOs: 47-96 and 178-179 are a set of overlapping sense and antisense oligonucleotides from erb-b, used for generating small dsRNA, as detailed more fully in Table 6.

SEQ ID NOs 97-167 are sense and antisense oligonucleotides used for generating small dsRNA, as detailed more fully in Table 3.

SEQ ID NOs: 168-177 are sense and antisense oligonucleotides used for generating small dsRNA, as detailed more fully in Table 5.

SEQ ID NOs: 180-311 are sense and antisense oligonucleotides used for generating small dsRNA, as detailed more fully in Table 8.

SEQ ID NO: 312 is the sequence of a portion of the *egfp* encoding sequence that includes Region B, illustrated to show alignment of the small dsRNAs show in Table 4.

SEQ ID NOs 313 and 314 are synthetic oligonucleotide primers used to pepare a human androgen rceptor gene specific probe.

DETAILED DESCRIPTION

I. Abbreviations

AR	androgen receptor
as	antisense
caspase	cysteine aspartate-specific proteases
CAT	chloramphenicol acetyl transferase
DEVD-AFC	Asp-Glu-Val-Asp-7-amino-4-trifluoromethylcoumarin

dsRNA	double stranded RNA
FACs	fluorescence activated cell sort analysis
FBS	fetal bovine serum.
FITC	fluoroscein
gapdh-1	glyceraldehyde 3-phosphate dehydrogenase-1
GFP	green fluorescence protein
HD	Huntington's disease (HD)
LacZ	β-galactosidase gene.
nt(s)	nucleotides(s)
MEFs	Mouse embryonic fibroblasts
PI	propidium iodide
PKR	dsRNA dependent protein kinase
polyQ	polyglutamine
PTGS	post-transcriptional gene silencing
Rh	Rhodamine
RNAi	RNA interference
s	sense
SBMA	spinobulbar muscular atrophy
siRNA	short interfering RNA
ssRNA	single stranded RNA

II. Terms

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes V*, published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

In order to facilitate review of the various embodiments of the invention, the following explanations of specific terms are provided:

Analog, derivative or mimetic: An analog is a molecule that differs in chemical structure from a parent compound, for example a homolog (differing by an increment in the chemical structure, such as a difference in the length of an alkyl chain), a molecular fragment, a structure that differs by one or more functional groups, a change in ionization. Structural analogs are often found using quantitative structure activity relationships (QSAR), with techniques such as those disclosed in Remington (*The Science and Practice of Pharmacology*, 19th Edition (1995), chapter 28). A derivative is a substance related to a base structure, and theoretically derivable from the base structure. A mimetic is a biomolecule that mimics the activity of another biologically active molecule. Biologically active molecules can include chemical structures that mimic the biological activities of a compound, for instance a native siRNA.

Anti-proliferative activity: An activity of a molecule, *e.g.*, a compound, which reduces proliferation of at least one cell type, but which may reduce the proliferation (either in absolute terms or in rate terms) of multiple different cell types (*e.g.*, different cell lines, different species, etc.). In specific embodiments, an anti-proliferative activity will be apparent against cells (either *in vitro* or *in*

vivo) that exhibit a hyper-proliferative condition, such as is characteristic of certain disorders or diseases.

In certain embodiments, an anti-proliferative activity can be an anti-tumor or anti-neoplastic activity of a compound. Such molecules will be useful to inhibit or prevent or reduce cellular proliferation or growth, *e.g.*, in a tumor, such as a malignant neoplasm.

Antisense, Sense, and Antigene: DNA has two strands, a 5' → 3' strand, referred to as the plus strand, and a 3' → 5' strand, referred to as the minus strand. Because RNA polymerase adds nucleic acids in a 5' → 3' direction, the minus strand of the DNA serves as the template for the RNA during transcription. Thus, the RNA formed will have a sequence complementary to the minus strand, and identical to the plus strand (except that U is substituted for T).

Antisense molecules are molecules that are specifically hybridizable or specifically complementary to either RNA or the plus strand of DNA. Sense molecules are molecules that are specifically hybridizable or specifically complementary to the minus strand of DNA. Antigene molecules are either antisense or sense molecules directed to a DNA target. An antisense RNA (asRNA) is a molecule of RNA complementary to a sense (encoding) nucleic acid molecule.

Array: An arrangement of molecules, particularly biological macromolecules (such as polypeptides or nucleic acids) in addressable locations on a substrate. The array may be regular (arranged in uniform rows and columns, for instance) or irregular. The number of addressable locations on the array can vary, for example from a few (such as three) to more than 50, 100, 200, 500, 1000, 10,000, or more. A "microarray" is an array that is miniaturized so as to require or benefit from microscopic examination, or other magnification, for its evaluation. Further miniaturization can be used to produce "nanoarrays."

Within an array, each arrayed molecule sample is addressable, in that its location can be reliably and consistently determined within the at least two dimensions of the array surface. In ordered arrays, the location of each molecule sample can be assigned to the sample at the time when it is spotted or otherwise applied onto the array surface, and a key may be provided in order to correlate each location with the appropriate target. Often, ordered arrays are arranged in a symmetrical grid pattern, but samples could be arranged in other patterns (*e.g.*, in radially distributed lines, spiral lines, or ordered clusters). Addressable arrays are computer readable, in that a computer can be programmed to correlate a particular address on the array with information (such as hybridization or binding data, including for instance signal intensity). In some examples of computer readable formats, the individual "spots" on the array surface will be arranged regularly in a pattern (*e.g.*, a Cartesian grid pattern) that can be correlated to address information by a computer.

The sample application "spot" on an array may assume many different shapes. Thus, though the term "spot" is used, it refers generally to a localized deposit of nucleic acid, and is not limited to a round or substantially round region. For instance, substantially square regions of mixture application can be used with arrays encompassed herein, as can be regions that are substantially rectangular (such as a slot blot-type application), or triangular, oval, or irregular. The shape of the array substrate itself

is also immaterial, though it is usually substantially flat and may be rectangular or square in general shape.

cDNA (complementary DNA): A piece of DNA lacking internal, non-coding segments (introns) and transcriptional regulatory sequences. cDNA may also contain untranslated regions (UTRs) that are responsible for translational control in the corresponding RNA molecule. cDNA is usually synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells or other samples.

Complementarity and Percentage Complementarity: Molecules with complementary nucleic acids form a stable duplex or triplex when the strands bind, or hybridize, to each other by forming Watson-Crick, Hoogsteen or reverse Hoogsteen base pairs. Stable binding occurs when an oligonucleotide remains detectably bound to a target nucleic acid sequence under the required conditions.

Complementarity is the degree to which bases in one nucleic acid strand base pair with the bases in a second nucleic acid strand. Complementarity is conveniently described by the percentage, *i.e.*, the proportion of nucleotides that form base pairs between two strands or within a specific region or domain of two strands. For example, if 10 nucleotides of a 15-nucleotide oligonucleotide form base pairs with a targeted region of a DNA molecule, that oligonucleotide is said to have 66.67% complementarity to the region of DNA targeted.

"Sufficient complementarity" means that a sufficient number of base pairs exist between the oligonucleotide and the target sequence to achieve detectable binding, and disrupt or reduce expression of the gene product(s) encoded by that target sequence. When expressed or measured by percentage of base pairs formed, the percentage complementarity that fulfills this goal can range from as little as about 50% complementarity to full, (100%) complementary. In some embodiments, sufficient complementarity is at least about 50%, about 75% complementarity, or at least about 90% or 95% complementarity. In particular embodiments, sufficient complementarity is 98% or 100% complementarity.

A thorough treatment of the qualitative and quantitative considerations involved in establishing binding conditions that allow one skilled in the art to design appropriate oligonucleotides for use under the desired conditions is provided by Beltz *et al.*, *Methods Enzymol* 100:266-285, 1983, and by Sambrook et al. (ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

Coupling: The term "coupling" refers to the chemical reaction of a molecule, such as a nucleotide, with a detectable molecule, such as a hapten or label (*e.g.*, a fluorophore). By way of example, coupling reactions may be reactions between a nucleophile (functional group) and an electrophile, *i.e.*, an electron poor reactive group. The coupling reaction may be facilitated by using an activating moiety to activate the electrophile to nucleophilic coupling. The activating group also usually is a leaving group. The nucleophile can be either on the nucleotide or on the detectable molecule, so long as the pair of reactants (nucleotide and detectable molecule) is capable of reacting with each other. Many embodiments have the nucleophile provided by the nucleotide.

Examples of reactions that may occur between the nucleophile and the electron poor reactive group include (in no particular order), but are not limited to, a Grignard reaction, a Wittig reaction, a condensation (such as an aldol condensation), a Mitsunobu reaction, formation of a Schiff base, and so forth.

Representative examples of nucleophilic functional groups include amines ($-NH_2$), -NHR (where R is aliphatic, *e.g.*, an alkyl group), alcohols (-OH), thiols (-SH), acido-acetates, alkyl lithium components, and so forth. Hydrogen-bearing compounds also can be deprotonated to facilitate the coupling reaction. Additional examples of functional groups will be apparent to one of ordinary skill in the art.

Representative examples of leaving groups include halides (including F, Cl, and I), sulfonates, phosphates, DCC, EDC, imidazole, DMAP, DMF/acid chloride, and so forth. Further leaving groups are listed, for instance, in U.S. Patent No. 5,268,486, and include isothiocyanate, isocyanate, monochlorotriazine, dichlortriazine, mono- or di-halogen substituted pyridine, mono- or di-halogen substituted diazine, maleimide, aziridine, sulfonyl halide, acid halide, hydroxysuccinimide ester, hydroxysulfosuccinimide ester, imido ester, hydrazine, azidonitrophenyl, azide, 3-(2-pridyl dithio)-proprionamide, glyoxal and aldehyde. Additional examples of leaving groups will be apparent to one of ordinary skill in the art.

Specific examples of coupling reactions between aminoallyl nucleotides and fluorophores and haptens are illustrated in Nimmakayalu *et al.* (*BioTechniques* 28:518-522, 2000).

It is specifically contemplated that a molecule may be coupled to the surface of an array, or to another molecule that has been in some way attached (*e.g.*, covalently) to the surface of an array.

Feature: An addressable spot/element containing a molecule or mixture of molecules on an array. Features may be created by printing the molecule(s), usually within some type of matrix, onto the array platform by a printing device, such as a quill like pen, or by a touch-less deposition system (see, *e.g.*, Harris *et al.*, *Nature Biotech.* 18:384-385, 2000). Alternatively, in some embodiments the feature can be made by in situ synthesis of the molecule on the array substrate.

Fluorophore: A chemical compound, which when excited by exposure to a particular wavelength of light, emits light (*i.e.*, fluoresces), for example at a different wavelength than that to which it was exposed. Fluorophores can be described in terms of their emission profile, or "color." Green fluorophores, for example Cy3, FITC, and Oregon Green, are characterized by their emission at wavelengths generally in the range of 515-540 λ. Red fluorophores, for example Texas Red, Cy5 and tetramethylrhodamine, are characterized by their emission at wavelengths generally in the range of 590-690 λ.

Encompassed by the term "fluorophore" are luminescent molecules, which are chemical compounds which do not require exposure to a particular wavelength of light to fluoresce; luminescent compounds naturally fluoresce. Therefore, the use of luminescent signals eliminates the need for an external source of electromagnetic radiation, such as a laser. An example of a luminescent molecule includes, but is not limited to, aequorin (Tsien, *Ann. Rev. Biochem.* 67:509, 1998).

Examples of fluorophores are provided in U.S. Patent No. 5,866,366. These include: 4-acetamido-4'-isothiocyanatostilbene-2,2'disulfonic acid, acridine and derivatives such as acridine and acridine isothiocyanate, 5-(2'-aminoethyl)aminonaphthalene-1-sulfonic acid (EDANS), 4-amino-N-[3-vinylsulfonyl)phenyl]naphthalimide-3,5 disulfonate (Lucifer Yellow VS), N-(4-anilino-1-naphthyl)maleimide, anthranilamide, Brilliant Yellow, coumarin and derivatives such as coumarin, 7-amino-4-methylcoumarin (AMC, Coumarin 120), 7-amino-4-trifluoromethylcouluarin (Coumaran 151); cyanosine; 4',6-diaminidino-2-phenylindole (DAPI); 5', 5"-dibromopyrogallol-sulfonephthalein (Bromopyrogallol Red); 7-diethylamino-3-(4'-isothiocyanatophenyl)-4-methylcoumarin; diethylenetriamine pentaacetate; 4,4'-diisothiocyanatodihydro-stilbene-2,2'-disulfonic acid; 4,4'-diisothiocyanatostilbene-2,2'-disulfonic acid; 5-[dimethylamino]naphthalene-1-sulfonyl chloride (DNS, dansyl chloride); 4-(4'-dimethylaminophenylazo)benzoic acid (DABCYL); 4-dimethylaminophenyl-azophenyl-4'-isothiocyanate (DABITC); eosin and derivatives such as eosin and eosin isothiocyanate; erythrosin and derivatives such as erythrosin B and erythrosin isothiocyanate; ethidium; fluorescein and derivatives such as 5-carboxyfluorescein (FAM), 5-(4,6-dichlorotriazin-2-yl)aminofluorescein (DTAF), 2'7'-dimethoxy-4'5'-dichloro-6-carboxyfluorescein (JOE), fluorescein, fluorescein isothiocyanate (FITC), and QFITC (XRITC); fluorescamine; IR144; IR1446; Malachite Green isothiocyanate; 4-methylumbelliferone; ortho cresolphthalein; nitrotyrosine; pararosaniline; Phenol Red; B-phycoerythrin; o-phthaldialdehyde; pyrene and derivatives such as pyrene, pyrene butyrate and succinimidyl 1-pyrene butyrate; Reactive Red 4 (Cibacron .RTM. Brilliant Red 3B-A); rhodamine and derivatives such as 6-carboxy-X-rhodamine (ROX), 6-carboxyrhodamine (R6G), lissamine rhodamine B sulfonyl chloride, rhodamine (Rhod), rhodamine B, rhodamine 123, rhodamine X isothiocyanate, sulforhodamine B, sulforhodamine 101 and sulfonyl chloride derivative of sulforhodamine 101 (Texas Red); N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA); tetramethyl rhodamine; tetramethyl rhodamine isothiocyanate (TRITC); riboflavin; rosolic acid and terbium chelate derivatives.

Other fluorophores include thiol-reactive europium chelates that emit at approximately 617 nm (Heyduk and Heyduk, *Analyt. Biochem.* 248:216-227, 1997; *J. Biol. Chem.* 274:3315-3322, 1999).

Still other fluorophores include cyanine, merocyanine, styryl, and oxonyl compounds, such as those disclosed in U.S. Patent Nos. 5,268,486; 5,486,616; 5,627,027; 5,569,587; and 5,569,766, and in published PCT patent application no. US98/00475, each of which is incorporated herein by reference. Specific examples of fluorophores disclosed in one or more of these patent documents include Cy3 and Cy5, for instance.

Other fluorophores include GFP, Lissamine™, diethylaminocoumarin, fluorescein chlorotriazinyl, naphthofluorescein, 4,7-dichlororhodamine and xanthene (as described in U.S. Patent No. 5,800,996 to Lee *et al.*, herein incorporated by reference) and derivatives thereof. Other fluorophores are known to those skilled in the art, for example those available from Molecular Probes (Eugene, OR).

Gene Silencing: Gene silencing refers to lack of (or reduction of) gene expression as a result of, though not limited to, effects at a genomic (DNA) level such as the generation of null mutant transgenic knock-out and chromatin re-structuring, as well as transcriptional, post-transcriptional gene silencing and translational blockade. Current evidence suggests that RNAi acts primarily at a post-transcriptional gene silencing, however, there is also evidence for gene silencing through genomic modification and translational blockade under specific circumstances. Antisense molecules primarily act through translational blockade, though there is also evidence for action at a transcriptional level as a result of RNA degradation.

Because RNAi exerts its effects at a transcriptional and post-transcriptional level, it is believed that RNAi can be used to specifically inhibit alternative transcripts from the same gene. This is important considering recent findings that while there are fewer genes within the human genome than previously thought, there is now an emphasis on investigating alternative transcripts from the same gene, which may account for protein complexity.

High throughput genomics: Application of genomic or genetic data or analysis techniques that use microarrays or other genomic technologies to rapidly identify large numbers of genes or proteins, or distinguish their structure, expression or function of genes or gene products from normal or abnormal cells or tissues.

Human cells: Cells obtained from a member of the species *Homo sapiens*. The cells can be obtained from any source, for example peripheral blood, urine, saliva, tissue biopsy, surgical specimen, amniocentesis samples and autopsy material. From these cells, genomic DNA, cDNA, mRNA, RNA, and/or protein can be isolated. Specific examples of human cells are cells derived from a neoplastic cell, for instance cancer cell lines.

Injectable composition: A pharmaceutically acceptable fluid composition including at least one active ingredient (for instance, a dsRNA). The active ingredient is usually dissolved or suspended in a physiologically acceptable carrier, and the composition can additionally include minor amounts of one or more non-toxic auxiliary substances, such as emulsifying agents, preservatives, and pH buffering agents and the like. Such injectable compositions that are useful for use with the provided nucleotides and proteins are conventional; appropriate formulations are well known in the art.

Interefering with or inhibiting (expression of a target gene): This phrase refers to the ability of a dsRNA or other molecule to measurably reduce the expression of a target gene. It contemplates reduction of the end-product of the gene, *e.g.*, the expression or function of the encoded protein, and thus includes reduction in the amount or longevity of of the mRNA transcript. It is understood that the phrase is relative, and does not require absolute suppression of the gene. Thus, in certain embodiments, interfering with or inhibiting gene expression of a target gene requires that, following application of the dsRNA, the gene is expressed at least 5% less than prior to application, at least 10% less, at least 15% less, at least 20% less, at least 25% less, or even more reduced. Thus, in some particular embodiments, application of a dsRNA reduces expression of the target gene by

about 30%, about 40%, about 50%, about 60%, or more. In specific examples, where the dsRNA is particularly effective, expression is reduced by 70%, 80%, 85%, 90%, 95%, or even more.

Nucleotide: "Nucleotide" includes, but is not limited to, a monomer that includes a base linked to a sugar, such as a pyrimidine, purine or synthetic analogs thereof, or a base linked to an amino acid, as in a peptide nucleic acid (PNA). A nucleotide is one monomer in an oligonucleotide/polynucleotide. A nucleotide sequence refers to the sequence of bases in an oligonucleotide/polynucleotide.

The major nucleotides of DNA are deoxyadenosine 5'-triphosphate (dATP or A), deoxyguanosine 5'-triphosphate (dGTP or G), deoxycytidine 5'-triphosphate (dCTP or C) and deoxythymidine 5'-triphosphate (dTTP or T). The major nucleotides of RNA are adenosine 5'-triphosphate (ATP or A), guanosine 5'-triphosphate (GTP or G), cytidine 5'-triphosphate (CTP or C) and uridine 5'-triphosphate (UTP or U). Inosine is also a base that can be integrated into DNA or RNA in a nucleotide (dITP or ITP, respectively).

Oligonucleotide: An oligonucleotide is a plurality of nucleotides joined by phosphodiester bonds, between about 6 and about 300 nucleotides in length. An oligonucleotide analog refers to compounds that function similarly to oligonucleotides but have non-naturally occurring portions. For example, oligonucleotide analogs can contain non-naturally occurring portions, such as altered sugar moieties or inter-sugar linkages, such as a phosphorothioate oligodeoxynucleotide. Functional analogs of naturally occurring polynucleotides can bind to RNA or DNA, and include peptide nucleic acid (PNA) molecules.

Parenteral: Administered outside of the intestine, *e.g.*, not via the alimentary tract. Generally, parenteral formulations are those that will be administered through any possible mode except ingestion. This term especially refers to injections, whether administered intravenously, intrathecally, intramuscularly, intraperitoneally, or subcutaneously, and various surface applications including intranasal, intradermal, and topical application, for instance.

Peptide Nucleic Acid (PNA): An oligonucleotide analog with a backbone comprised of monomers coupled by amide (peptide) bonds, such as amino acid monomers joined by peptide bonds.

Pharmaceutically acceptable carriers: The pharmaceutically acceptable carriers useful with compositions provided herein are conventional. Martin, *Remington's Pharmaceutical Sciences*, published by Mack Publishing Co., Easton, PA, 19th Edition, 1995, describes compositions and formulations suitable for pharmaceutical delivery of the nucleotides and proteins herein disclosed.

In general, the nature of the carrier will depend on the particular mode of administration being employed. For instance, parenteral formulations usually comprise injectable fluids that include pharmaceutically and physiologically acceptable fluids such as water, physiological saline, balanced salt solutions, aqueous dextrose, glycerol or the like as a vehicle. For solid compositions (*e.g.*, powder, pill, tablet, or capsule forms), conventional non-toxic solid carriers can include, for example, pharmaceutical grades of mannitol, lactose, starch, or magnesium stearate. In addition to biologically-neutral carriers, pharmaceutical compositions to be administered can contain minor

amounts of non-toxic auxiliary substances, such as wetting or emulsifying agents, preservatives, and pH buffering agents and the like, for example sodium acetate or sorbitan monolaurate.

Polymerization: Synthesis of a nucleic acid chain (oligonucleotide or polynucleotide) by adding nucleotides to the hydroxyl group at the 3'-end of a pre-existing RNA or DNA primer using a pre-existing DNA strand as the template. Polymerization usually is mediated by an enzyme such as a DNA or RNA polymerase. Specific examples of polymerases include the large proteolytic fragment of the DNA polymerase I of the bacterium *E. coli* (usually referred to as Kleenex polymerase), *E. coli* DNA polymerase I, and bacteriophage T7 DNA polymerase. Polymerization of a DNA strand complementary to an RNA template (*e.g.*, a cDNA complementary to a mRNA) can be carried out using reverse transcriptase (in a reverse transcription reaction).

For *in vitro* polymerization reactions, it is necessary to provide to the assay mixture an amount of required cofactors such as M^{++}, and dATP, dCTP, dGTP, dTTP, ATP, CTP, GTP, UTP, or other nucleoside triphosphates, in sufficient quantity to support the degree of polymerization desired. The amounts of deoxyribonucleotide triphosphates substrates required for polymerizing reactions are well known to those of ordinary skill in the art. Nucleoside triphosphate analogues or modified nucleoside triphosphates can be substituted or added to those specified above.

Primer: Primers are relatively short nucleic acid molecules, usually DNA oligonucleotides six nucleotides or more in length. Primers can be annealed to a complementary target DNA strand ("priming") by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then the primer extended along the target DNA strand by a nucleic acid polymerase enzyme. Pairs of primers can be used for amplification of a nucleic acid sequence, *e.g.*, by nucleic-acid amplification methods known in to those of ordinary skill in the art.

A primer is usually single stranded, which may increase the efficiency of its annealing to a template and subsequent polymerization. However, primers also may be double stranded. A double stranded primer can be treated to separate the two strands, for instance before being used to prime a polymerization reaction (see for example, *Nucleic Acid Hybridization. A Practical Approach.* Hames and Higgins, eds., IRL Press, Washington, 1985). By way of example, a double stranded primer can be heated to about 90°-100° C. for about 1 to 10 minutes.

Recombinant: A recombinant nucleic acid is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination can be accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, *e.g.*, by genetic engineering techniques.

RNA: A typically linear polymer of ribonucleic acid monomers, linked by phosphodiester bonds. Naturally occurring RNA molecules fall into three general classes, messenger (mRNA, which encodes proteins), ribosomal (rRNA, components of ribosomes), and transfer (tRNA, molecules responsible for transferring amino acid monomers to the ribosome during protein synthesis). Messenger RNA includes heteronuclear (hnRNA) and membrane-associated polysomal RNA

(attached to the rough endoplasmic reticulum). Total RNA refers to a heterogeneous mixture of all types of RNA molecules.

Specific binding agent: An agent that binds substantially only to a defined target. Thus a protein-specific binding agent binds substantially only the specified protein.

Anti-target antibodies may be produced using standard procedures described in a number of texts, including Harlow and Lane (*Antibodies, A Laboratory Manual*, CSHL, New York, 1988). The determination that a particular agent binds substantially only to the specified target molecule (*e.g.*, a protein) may readily be made by using or adapting routine procedures. One suitable *in vitro* assay makes use of the Western blotting procedure (described in many standard texts, including Harlow and Lane (*Antibodies, A Laboratory Manual*, CSHL, New York, 1988)). Western blotting may be used to determine that a given binding agent binds substantially only to the desired target molecule.

Shorter fragments of antibodies can also serve as specific binding agents. For instance, Fabs, Fvs, and single-chain Fvs (SCFvs) that bind to a specified protein would be specific binding agents. These antibody fragments are described as follows: (1) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain; (2) Fab', the fragment of an antibody molecule obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule; (3) (Fab')2, the fragment of the antibody obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; (4) F(ab')2, a dimer of two Fab' fragments held together by two disulfide bonds; (5) Fv, a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and (6) single chain antibody ("SCA"), a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule. Methods of making these fragments are routine.

Silencing agent or molecule: A specific molecule, which can exert an influence on a cell in a sequence-specific manner to reduce or silence the expression or function of a target, such as a target gene or protein. Examples of silence agents include nucleic acid molecules such as naturally occurring small interfering RNAs (siRNAs), naturally occurring or synthetically generated dsRNAs and antisense sequences (including antisense oligonucleotides, hairpin structures, and antisense expression vectors), as well as constructs that code for dominant negative peptides.

Small dsRNAs: Double stranded RNAs (dsRNAs) that can induce gene-specific inhibition of expression in invertebrate and vertebrate species are provided. These RNAs are suitable for interference or inhibition of expression of a target gene and comprise double stranded RNAs of about 15 to about 40 nucleotides containing a 3' and/or 5' overhang on each strand having a length of 0- to about 5-nucleotides, wherein the sequence of the double stranded RNAs is substantially identical to a portion of a mRNA or transcript of the target gene for which interference or inhibition of expression

is desired. The double stranded RNAs can be formed from complementary ssRNAs or from a single stranded RNA that forms a hairpin or from expression from a DNA vector.

In addition to native RNA molecules, RNA suitable for inhibiting or interfering with the expression of a target sequence include RNA derivatives and analogs. For example, a non-natural linkage between nucleotide residues can be used, such as a phosphorothioate linkage. The RNA strand can be derivatized with a reactive functional group or a reporter group, such as a fluorophore. Particularly useful derivatives are modified at a terminus or termini of an RNA strand, typically the 3' terminus of the sense strand. For example, the 2'-hydroxyl at the 3' terminus can be readily and selectively derivatized with a variety of groups.

Other useful RNA derivatives incorporate nucleotides having modified carbohydrate moieties, such as 2'-O-alkylated residues or 2'-deoxy-2'-halogenated derivatives. Particular examples of such carbohydrate moieties include 2'-O-methyl ribosyl derivatives and 2'-O-fluoro ribosyl derivatives.

The RNA bases may also be modified. Any modified base useful for inhibiting or interfering with the expression of a target sequence can be used. For example, halogenated bases, such as 5-bromouracil and 5-iodouracil can be incorporated. The bases can also be alkylated, for example, 7-methylguanosine can be incorporated in place of a guanosine residue. Non-natural bases that yield successful inhibition can also be incorporated.

Additional specific modifications are discussed in the text; some are illustrated also in Figures 11 and 12.

Unless otherwise explained, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The singular terms "a," "an," and "the" include plural referents unless context clearly indicates otherwise. Similarly, the word "or" is intended to include "and" unless the context clearly indicates otherwise. Hence "comprising A or B" means include A, or B, or A and B. It is further to be understood that all base sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are approximate, and are provided for description. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

III. Overview of Several Embodiments

One embodiment is an RNA for interference or inhibition of expression of a target gene, which RNA includes double-stranded RNA of about 15 to about 40 nucleotides in length and a 3' or 5' overhang having a length of 0-nucleotide to 5-nucleotides on each strand, wherein the sequence of

the double stranded RNA is substantially identical to (*e.g.*, no more than 30% different from) a portion of a mRNA or transcript of the target gene. In particular examples, the double stranded RNA contains about 19 to about 25 nucleotides, for instance 20, 21, or 22 nucleotides.

Regarding the overhang on the double-stranded RNA, the length of the overhang is independent between the two strands, in that the length of one overhang is not dependent on the length of the overhang on other strand. In specific examples, the length of the 3' or 5' overhang is 0-nucleotide on at least one strand, and in some cases it is 0-nucleotide on both strands (thus, a blunt dsRNA). In other examples, the length of the 3' or 5' overhang is 1-nucleotide to 5-nucleotides on at least one strand. More particularly, in some examples the length of the 3' or 5' overhang is 2-nucleotides on at least one strand, or 2-nucleotides on both strands. In particular examples, the dsRNA molecule has 3' overhangs of 2-nucleotides on both strands.

Thus, in one particular provided RNA embodiment, the double-stranded RNA contains 20, 21, or 22 nucleotides, and the length of the 3' overhang is 2-nucleotides on both strands.

In embodiments of the RNAs provided herein, the double-stranded RNA contains about 40-60% adenine+uracil (AU) and about 60-40% guanine+cytosine (GC). More particularly, in specific examples the double-stranded RNA contains about 50% AU and about 50% GC.

Also described herein are RNAs that further include at least one modified ribonucleotide, for instance in the sense strand of the double-stranded RNA. In particular examples, the modified ribonucleotide is in the 3' overhang of at least one strand, or more particularly in the 3' overhang of the sense strand. It is particularly contemplated that examples of modified ribonucleotides include ribonucleotides that include a detectable label (for instance, a fluorophore, such as rhodamine or FITC), a thiophosphate nucleotide analog, a deoxynucleotide (considered modified because the base molecule is ribonucleic acid), a 2'-fluorouracil, a 2'-aminouracil, a 2'-aminocytidine, a 4-thiouracil, a 5-bromouracil, a 5-iodouracil, a 5-(3-aminoallyl)-uracil, an inosine, or a 2'O-Me-nucleotide analog.

The target gene in some examples is an endogenous gene in a cell, for instance a cell that is *in vivo* in an organism. By way of example, the organism may be an invertebrate animal (*e.g.*, a nematode such as *C. elegans* or a fly, such as a *Drosophila*)or a vertebrate animal (such as a mammal, for instance a mouse, a monkey, or a human).

Particular examples of RNAs for interference or inhibition of expression of a target gene include double-stranded RNAs that include a sequence as shown in any one of SEQ ID NOs: 1 through 311.

Another embodiment is a method of interfering with or inhibiting expression of a target gene in a cell, the method comprising exposing (once or repeatedly) the cell to an effective amount of any of the RNAs provided herein. In examples of this embodiment, the cell is an animal cell (*in vivo* or *in vitro*), for instance a mammalian cell such as a human cell (*in vivo* or *in vitro*). Particular examples of such methods are genetic methods for treating an animal.

Yet another embodiment is a gene silencing array, which array includes a substantially flat substrate; and addressably arrayed thereon, a plurality of different double-stranded RNAs, each having a length of about 15 to 40 nucleotides, a 3' or 5' overhang having a length of 0-nucleotides to

5-nucleotides on each strand, and wherein the sequence of each double-stranded RNA is substantially identical to a portion of a mRNA or transcript of a target gene. In examples of such gene silencing arrays, the sequence of at least 50% of the double-stranded RNAs on the array is substantially identical to portions of a mRNA or transcript of a single target gene. In still other examples, the sequence of at least 70% of the double-stranded RNAs is substantially identical to portions of a mRNA or transcript of a single target gene. By way of example, in specific gene silencing arrays the target gene(s) is *Erb-B2*, *APPBP2*, *BMP7*, *CCND1*, *CRYM*, *ER1*, *FKBP5*, *FLJ20940*, *GRB7*, *HOXB7*, *LMO4*, *MGC9753*, *MLN64*, *MYBL2*, *MYC*, *NBS1*, *NCOA3*, *PIP5K2B*, *PNMT*, *PPARBP*, *PPM1D*, *RAD51C*, *RAE1*, *RPS6K. S100P*, *TBX2*, *TMEPAI*, *TRIM37*, *TXNIP*, or *ZNF217*.

Examples of gene silencing arrays include at least 50 different double stranded RNAs, for instance at least 100, at least 200, at least 500, or at least 800 different double-stranded RNAs. Particular contemplated gene silencing arrays include more than 1000 different double-stranded RNAs. The arrays may include double-stranded RNAs directed to a single target gene, or more often will contain double-stranded RNAs to a plurality of target genes.

In particular embodiments, the substantially flat substrate of the gene silencing array is a microscope slide, a chamber slide, a culture plate, or a 96-well microtiter plate.

Another embodiment is an array-based method of assessing a phenotypic effect of a double-stranded RNA on a target gene, which method involves addressably arraying the double-stranded RNA on a substrate, wherein the double-stranded RNA has a length of about 15 to 40 nucleotides, a 3' or 5' overhang having a length of 0-nucleotides to 5-nucleotides on each strand, and wherein the sequence of the double-stranded RNA is substantially identical to a portion of a mRNA or transcript of the target gene; overlaying the substrate with cells; culturing the cells under conditions sufficient to enable at least a portion of the double-stranded RNA to be taken up into at least one of the cells; and assessing at least one phenotypic characteristic of the at least one of the cells, wherein a change in the phenotypic characteristic in a cell that has taken up a double-stranded RNA compared to a cell that has not taken up the double-stranded RNA, or that has taken up a different double-stranded RNA, is indicative of a phenotypic effect.

In examples of such methods, the cells are *in vitro* cultured animal cells, such as mammalian (*e.g.*, human) cells. In particular examples, the cells are cancer cells.

In other examples, the target gene in the array-based method is *Erb-B2*, *APPBP2*, *BMP7*, *CCND1*, *CRYM*, *ER1*, *FKBP5*, *FLJ20940*, *GRB7*, *HOXB7*, *LMO4*, *MGC9753*, *MLN64*, *MYBL2*, *MYC*, *NBS1*, *NCOA3*, *PIP5K2B*, *PNMT*, *PPARBP*, *PPM1D*, *RAD51C*, *RAE1*, *RPS6K. S100P*, *TBX2*, *TMEPAI*, *TRIM37*, *TXNIP*, or *ZNF217*.

Also provided are methods of assessing a phenotypic effect of a plurality of double-stranded RNAs on the target gene, further comprising addressably arraying the plurality of double-stranded RNAs on the substrate (*e.g.*, a microscope slide, a chamber slide, a culture plate, or a 96-well microtiter plate). In examples of such methods, the sequence of at least one of the plurality of double-stranded RNAs is substantially identical to a portion of a mRNA or transcript of a first target gene and the sequence of at least one of the plurality of the double-stranded RNAs is substantially

identical to a portion of a mRNA or transcript of a second target gene, and wherein the method is a method of assessing a phenotypic effect of the plurality of double-stranded RNAs on a plurality of target genes. In certain embodiments at least two of the plurality of target genes are selected from the group consisting of *Erb-B2*, *APPBP2*, *BMP7*, *CCND1*, *CRYM*, *ER1*, *FKBP5*, *FLJ20940*, *GRB7*, *HOXB7*, *LMO4*, *MGC9753*, *MLN64*, *MYBL2*, *MYC*, *NBS1*, *NCOA3*, *PIP5K2B*, *PNMT*, *PPARBP*, *PPM1D*, *RAD51C*, *RAE1*, *RPS6K*. *S100P*, *TBX2*, *TMEPAI*, *TRIM37*, *TXNIP*, and *ZNF217*.

Examples of the methods of assessing a phenotypic effect of one or a plurality of double-stranded RNAs on a target gene are provided wherein the phenotypic effect is an anti-cancer phenotypic effect. For instance, the anti-cancer phenotypic effect in some cases includes an anti-proliferative effect, a pro-apoptotic effect or a combination thereof.

Another embodiment is a method of validating a gene as a potential drug target for a disease or condition, which method involves assaying a plurality of dsRNAs (for instance, in an array-based analysis), each having a sequence substantially identical to a portion of a mRNA or transcript of the gene, for the ability to generate a desired phenotype in a cell, wherein the phenotype is a phenotype related to the disease or condition; and selecting at least one dsRNA that generates the desired phenotype. In examples of this method, the gene is a cancer-associated gene and the desired phenotype of the cell comprises an anti-cancer phenotype (for instance, including inhibition of proliferation, stimulation of apoptosis, or a combination thereof).

Still another embodiment is a method of selecting an optimized sequence of a double-stranded RNA for interference with or inhibition of expression of a target gene in a cell, which method involves selecting one or more sub-sequence from within the sequence of the target gene that satisfy a first criterion, where the first criterion is that each sub-sequence is about 15 to about 40 nucleotides long (or in some examples, about 19-23 nucleotides long); selecting one or more sub-sequence from within the sequence of the target gene that satisfy a second criterion, where the second criterion is that each sub-sequence has a AU:GC content percent of about 40:60 to 60:40 (or in some embodiments, about 50:50); selecting one or more sub-sequence from within the sequence of the target gene that satisfy a third criterion, where the third criterion is that each sub-sequence is unique in comparison to a gene database from the same organism as the cell; and selecting at least one sub-sequence that satisfies the first, second, and third criteria, which sub sequence is an optimized sequence of the double-stranded RNA for interference with or inhibition of expression of the target gene in the cell.

In particular examples of this method, selecting one or more sub-sequences to satisfy the first, second, and third criteria occur concurrently. In other examples, selecting one or more sub-sequences to satisfy the first, second, and third criteria occur sequentially.

Finally there are provided a number short double-stranded RNAs effective for interfering with or inhibiting expression of a target gene, which RNAs include a sequence as shown in any one of SEQ ID NOs: 1 through 311.

IV. Inhibition of Gene Expression by Small Double Stranded RNAs

The present disclosure provides RNA suitable for interference or inhibition of expression of a target gene, which RNA includes double stranded RNA of about 15 to about 40 nucleotides containing a 0-nucleotide to 5-nucleotide 3' and/or 5' overhang on each strand. The sequence of the RNA is substantially identical to a portion of a mRNA or transcript of a target gene for which interference or inhibition of expression is desired. For purposes of this disclosure, a sequence of the RNA "substantially identical" to a specific portion of the mRNA or transcript of the target gene for which interference or inhibition of expression is desired differs by no more than about 30 percent, and in some embodiments no more than about 10 percent, from the specific portion of the mRNA or transcript of the target gene. In particular embodiments, the sequence of the RNA is exactly identical to a specific portion of the mRNA or transcript of the target gene.

The present disclosure provides a method for treating animal cells by interfering or inhibiting expression of a target gene in the animal cells, said method comprising exposing the animal cells to an effective amount of RNA suitable for interfering or inhibiting expression of the target gene, wherein the RNA comprises double stranded RNA of about 15 to about 40 nucleotides containing a 0-nucleotide to 5-nucleotide 3' and/or 5' overhang on each strand, wherein the sequence of the RNA is substantially identical to a portion of a mRNA or transcript of the target gene. The present method is suitable, for instance, for the treatment of mammal cells, more especially for the treatment of human cells *in vivo* and *in vitro*.

The present disclosure also provides a genetic method for treating an animal by interfering or inhibiting expression of a target gene in the animal, said method comprising administering to the animal an effective amount of RNA suitable for interfering or inhibiting expression of the target gene, wherein the RNA comprises double stranded RNA of about 15 to about 40 nucleotides containing 0-nucleotide to 5-nucleotide 3' and/or 5' overhang on each strand, wherein the sequence of the RNA is substantially identical to a portion of a mRNA or transcript of the target gene. The present disclosure is suitable, for instance, for the treatment of mammals, more especially for the treatment of humans.

The double stranded RNAs of this disclosure may have unmodified or modified backbones and/or component nucleosides. Such modifications include, but are not limited to, -thio, 2'-fluro 2'-amino, 2'-deoxy, 4-thio, 5-bromo, 5-iodo and 5-(3-aminoallyl) derivatives of ribonucleosides.

A consistent observation of PTGS and RNAi in several species has been the detection of small dsRNAs (~21-25 nt) derived from the triggering dsRNA, which have been referred to as small interfering RNAs. These small dsRNAs have been observed irrespective of whether the initiating dsRNA is delivered directly, derived from a viral RNA, or produced from a transgene (Hamilton *et al.*, *Science* 286, 950, 1999; Zamore *et al.*, *Cell* 101, 25, 2000; Hammond *et al.*, *Nature* 404, 293, 2000; Yang *et al.*, *Curr. Biol.* 10, 1191, 2000; Parrish *et al.*, *Mol. Cell* 6, 1077, 2000; Elbashir *et al.*, *Genes Dev.* 15, 188, 2001). These findings and further biochemical analysis (Bernstein *et al.*, *Nature* 409, 363, 2001) have suggested that the generation of these small dsRNAs represents a critical step in the RNAi/PTGS mechanism. Presented herein is evidence that these small dsRNAs can have direct effects on gene expression in *C. elegans* and mammalian cell culture *in vivo*.

The results in mammalian cells are particularly striking, in that previous attempts to assay RNAi effects in vertebrate somatic cells have encountered effects that were predominately gene-nonspecific (Tuschl *et al.*, *Genes Dev.* 13, 3191, 1999; Caplen *et al.*, *Gene* 252, 95, 2000; Oates *et al.*, *Dev. Biol.* 224, 20, 2000; Zhao *et al.*, *Dev. Biol.* 229, 215, 2001). Although not wishing to be limited by theory, it is proposed herein that the small size of the double stranded RNAs avoids the induction of the non-specific responses of mammalian cells to dsRNA.

Several models have been put forward to explain RNAi, in particular the mechanisms by which the cleavage-derived small dsRNAs or siRNAs interact with the target mRNA and thus facilitate its degradation (Hamilton *et al.*, *Science* 286, 950, 1999; Zamore *et al.*, *Cell* 101, 25, 2000; Hammond *et al.*, *Nature* 404, 293, 2000; Yang *et al.*, *Curr. Biol.* 10, 1191, 2000; Elbashir *et al.*, *Genes Dev.* 15, 188, 2001; Bass *Cell* 101, 235, 2000). It has been proposed that the cleavage derived small dsRNAs or siRNAs act as a guide for the enzymatic complex required for the sequence specific cleavage of the target mRNA. Evidence for this includes cleavage of the target mRNA at regular intervals of ~21-23 nts in the region corresponding to the input dsRNA (Zamore *et al.*, *Cell* 101, 25, 2000), with the exact cleavage sites corresponding to the middle of sequences covered by individual 21 or 22 nt small dsRNAS or siRNAs (Elbashir *et al.*, *Genes Dev.* 15, 188, 2001). Although mammals and lower organisms appear to share dsRNA-triggered responses that involve a related intermediate (small dsRNAs), it is likely that there will be differences as well as similarities in the underlying mechanism.

Several of the proteins shown to play key roles in RNA-triggered gene silencing in plants and invertebrates share homology with potential coding regions from the human or other vertebrate genomes. These include putative RNA-dependent polymerases (RdRp) (Cogoni *et al.*, *Nature* 399, 166, 1999; Smardon *et al.*, *Curr Biol* 10, 169, 2000; Dalmay *et al.*, *Cell* 101, 543, 2000; Mourrain *et al.*, *Cell* 101, 533, 2000), the RDE-1/Argonaute family (Tabara *et al.*, *Cell* 99, 123, 1999), and a variety of putative helicases and nucleases (Ketting *et al.*, *Cell* 99, 133, 1999; Bernstein *et al.*, *Nature* 409, 363, 2001; Wu-Scharf *et al.*, *Science* 290, 1159, 2000; Dernburg *et al.*, *Genes Dev.* 14, 1578, 2000; Domeier *et al.*, *Science* 289, 1928, 2000). Mammalian homologs of the RNAi associated Drosophila RNase III have been identified (Matsuda *et al.*, *Biochim. Biophys. Acta* 1490:163-169, 2000; Wu *et al.*, *J. Biol. Chem.* 275:36957-36965, 2000). One of these putative RNases has been shown to generate small dsRNA molecules of approximately 22 nucleotides from larger dsRNAs (Bernstein *et al.*, *Nature* 409:363, 2001). However, even in invertebrate systems, the precise role of these factors in RNAi remains to be elucidated. Since factors from each of these homology classes have identified roles in normal physiology and development (*i.e.*, beyond genome surveillance), a full analysis of the reaction mechanisms in the different biological systems may be needed before a clear picture of the commonality between RNAi in these different systems will emerge.

Because of the efficacy and ease with which RNAi can be induced, RNAi has been rapidly exploited in *C. elegans* and *Drosophila* as a reverse genetics tool (Kuwabara *et al.*, *Parasitol Today* 16:347-349, 2000). Currently, the principal method used to reduce gene expression in mammalian cells utilizes antisense sequences in the form of single stranded oligonucleotides and transcripts. It is

believed that the interaction of antisense sequences with mRNA through Watson-Crick base-pairing leads to a decrease in gene expression via several possible mechanisms, such as the activation of RNase H which cleaves RNA/DNA duplexes and the inhibition of RNA processing and/or translational blockade (Crooke, *Biochim. Biophys. Acta* 1489:31-44, 1999).

Several issues have limited wider use of antisense technology. Problems have included a lack of suitable target sequences within a given mRNA due to RNA secondary folding, which necessitates screening of multiple antisense sequences to identify those that mediate the greatest level of inhibition, and inefficient delivery of antisense molecules *in vitro* and *in vivo*. All of the small dsRNAs that were examined produced specific inhibition of gene expression. The small dsRNAs appear to be very stable and thus may not require the extensive chemical modifications that ssRNA antisense oligonucleotide require to enhance the *in vivo* half-life.

V. Cell-Based Gene Silencing Microarrays

Since over-expressing transgenes may not produce physiologically relevant changes in endpoints, gene knockouts or other methods of gene silencing provide critically important information for characterizing gene function. Currently, large-scale somatic gene knockout studies have been difficult to perform in mammalian cells. Provided herein and in co-pending U.S. Provisional Application 60/370,970, are gene-silencing microarray technologies for high-throughput cellular analysis of gene function and demonstrated its capabilities in measuring the effects of transient gene silencing. The system is flexible and is expandable for functional characterization of genes on a genome scale through the parallel analysis of thousands of gene silencing events on a single microscope slide.

Provided herein are methods of adapting the transfected cell microarray platform for the highly parallel analysis of the effects of gene silencing in living cells, particularly using dsRNA to induce sequence-specific inhibition of RNA expression.

To address the challenges of accurately measuring the cellular and molecular consequences that result from gene silencing in a particular feature on an array, in certain embodiments analysis tools for resolution of multiparametric single cell resolution data extraction from microscopic images have also been developed.

Methods described herein offer a flexible system for the quantitative assessment of the causal effects and consequences of specific gene silencing. RNAi microarrays can therefore be used for a variety of applications from screening genes for a particular functional characteristic to more comprehensive multidimensional analysis revealing contextual networks of causal gene relationships.

The methods provided herein can be used with any type of cell that can be cultured in a monolayer (or approximate monolayer) on the surface of the array. These include, for instance, invertebrate and vertebrate cells, including mammalian cells such as human cells. Specifically contemplated herein are methods using immortalized mammalian cells, including for instance cells from various cancer cell lines. Individual cell lines include, for instance, the breast cancer cell line SKBR3 and the prostate cancer cell line PC3.

It is not essential that the transfection efficiency be particularly high for cells used in methods provided herein. In particular, cell-by-cell analysis of the gene silencing arrays can be used to correct for low transfection efficiency, as described more fully below.

Cells will normally form attached monolayers with a confluency of approximately 30% to 100%. Cells are generally grown under standard growth conditions, usually at 37° C and in the presence of carbon dioxide.

Essentially any cell type that is adherent can be utilized in gene silencing microarrays. Preferably, the cells should be growing in a monolayer. The cells can be plated on arrays created on microscope slides or on arrays printed on the surface of a cell culture flask/Petri dish. Cells can be grown in normal cell culture conditions, or challenged by incubation with physical or chemical stimuli at the same time as they are subjected to multiple "reverse transfection" procedures. To facilitate the latter approach, a replicate or a different microarray can be printed in the multiple wells of microtiter plates. See also U.S. Patent No. 5,545,531 (incorporated herein by reference), for another method of generating replicate microarrays that are available for multiple analyses.

If the cells are not adherent or growing as a monolayer, the surface of the array format can be coated to promote cell adhesion or covered in a matrix that retains cells *in situ*.

Alternatively, cells that are not adherent can be applied to the surface of the arrays in a suspension, such as a gel-based suspension similar to those used for producing "tissue array" samples from suspended cells or cells of non-solid tumors. See, for instance, methods disclosed in PCT/US99/04001 and PCT/US99/04000.

Gene silencing agents placed on the microarrays are taken up using, though not limited to, standard non-viral transfection methods including cationic lipid mediated nucleic acid transfer, cationic-polymer mediated nucleic acid transfer, calcium-phosphate precipitation and electroporation.

Gene silencing arrays can be produced and analyzed as described, for instance, in co-pending U.S. patent application 60/370,970, filed April 8, 2002, which is incorporated herein in its entirety.

VI. Methods of Using Cell-Based Gene Silencing Microarrays

The cell-based gene-silencing microarrays described herein can be used in any analysis for which it is beneficial to be able to knock out or reduce the expression of a gene or gene product. For instance, they can be used to examine the effects (on a number of endpoints) caused by gene silencing of known or unknown genes, gene fragments, and so forth.

In addition, provided gene silencing microarrays can be used to rapidly and efficiently examine the effectiveness of gene silencing molecules, such as antisense or dsRNA molecule. Thus, specific contemplated gene silencing arrays include a plurality of gene silencing agents directed to the same target gene and are used for characterization of these agents (see, *e.g.*, Example 3). One example of such an array includes, for instance, a series of (overlapping or non-overlapping) putative dsRNAs corresponding to a single target gene, such as a gene known to be involved or implicated in a disease or disease progression, for instance as a tumor-related gene. By applying cells to such an

array and examining the effectiveness of gene silencing at individual features (using one or more endpoints), specific highly effective gene silencing agents can be selected. Representative arrays of this type will include not only a plurality of one type of gene silencing agent directed to a specific target, but may also include a plurality of each of various different types of gene silencing agents. Thus, provided herein are specific arrays useful in identifying optimal gene silencing agents for specific target genes. Parallel analysis of such testing arrays, examining different endpoints, provides a particularly effective method for determining the effectiveness of the selected gene silencing agent(s).

Examples of such arrays are arrays directed towards dsRNAs that target the Erb-B2 gene. Specific examples of these arrays will include one or more of the dsRNAs shown in Tables 1 and 2.

Gene silence arrays can also be used to investigate drug targets, which have been identified (through any applicable method, including through the use of a gene silencing array). For instance, take as an example a gene has been identified as a drug target in a cDNA microarray analysis (or other analysis) because it is up-regulated in a particular disease or condition. A gene silencing array including one or more gene silencing agents directed to that target can be used to examine whether inhibition of that drug target is effective in changing an endpoint that is associated with that disease or condition.

Gene silencing agents identified as having an effect on an endpoint can themselves be selected as potential drugs for further examination.

VII. Using Identified Gene Silencing Agents

Methods are provided herein for the identification of specific dsRNAs that are effective, or particularly effective, at gene silencing of specific target molecules.

In particular, there are provided herein specific dsRNAs useful for gene silencing of the transgenes Green Fluorescent Protein (*egfp*), chloramphenicol acetyl transferase (*cat*) and ß-galactosidase (*LacZ*), the *C. elegans unc22* gene and the mammalian genes *Erb-B2*, *APPBP2*, *BMP7*, *CCND1*, *CRYM*, *ER1*, *FKBP5*, *FLJ20940*, *GRB7*, *HOXB7*, *LMO4*, *MGC9753*, *MLN64*, *MYBL2*, *MYC*, *NBS1*, *NCOA3*, *PIP5K2B*, *PNMT*, *PPARBP*, *PPM1D*, *RAD51C*, *RAE1*, *RPS6K. S100P*, *TBX2*, *TMEPAI*, *TRIM37*, *TXNIP*, and *ZNF217*. Many of these gene targets are very important drug targets , since it is known that they are involved in mediating cancers, for instance specific breast or prostate cancers.

With the provision herein of specific silencing dsRNAs (including the *Erb-B2* dsRNAs listed in Table 6 and others discussed in the Examples), methods of treating hyper-proliferative diseases associated with over-expression of *Erb-B2* and other cancer-associated genes (such as *APPBP2*, *BMP7*, *CCND1*, *CRYM*, *ER1*, *FKBP5*, *FLJ20940*, *GRB7*, *HOXB7*, *LMO4*, *MGC9753*, *MLN64*, *MYBL2*, *MYC*, *NBS1*, *NCOA3*, *PIP5K2B*, *PNMT*, *PPARBP*, *PPM1D*, *RAD51C*, *RAE1*, *RPS6K. S100P*, *TBX2*, *TMEPAI*, *TRIM37*, *TXNIP*, and/or *ZNF217*) are now enabled.

VIII. Methods of Treatment

The results presented herein indicate that small dsRNAs are useful for triggering RNAi-like responses that can be utilized as both functional genomics and therapeutic tools. Thus, the present disclosure includes methods of using dsRNAs as a treatment for disease. In a particular embodiment, the dsRNA used is a dsRNA corresponding to the *Erb-B2* gene, and as such the treatment is a treatment of a hyper-proliferative disease or disorder, such as breast cancer, in a subject. In other methods, another cancer-associated gene is implicated (for instance, *APPBP2*, *BMP7*, *CCND1*, *CRYM*, *ER1*, *FKBP5*, *FLJ20940*, *GRB7*, *HOXB7*, *LMO4*, *MGC9753*, *MLN64*, *MYBL2*, *MYC*, *NBS1*, *NCOA3*, *PIP5K2B*, *PNMT*, *PPARBP*, *PPM1D*, *RAD51C*, *RAE1*, *RPS6K. S100P*, *TBX2*, *TMEPAI*, *TRIM37*, *TXNIP*, and/or ZNF217), and a dsRNA to that cancer-associated gene is used.

The method includes administering a dsRNA, or more than one dsRNA, or a combination of a dsRNA (or more than one) and one or more other pharmaceutical agents, to the subject in a pharmaceutically compatible carrier and in an amount effective to inhibit the development or progression of a disease. Although the treatment can be used prophylactically in any patient in a demographic group at significant risk for such diseases, subjects can also be selected using more specific criteria, such as a definitive diagnosis of the disease/condition or identification of one or more factors that increase the likelihood of developing such disease (*e.g.*, a genetic, environmental, or lifestyle factor).

Various delivery systems are known and can be used to administer the small dsRNAs as therapeutics. Such systems include, for example, encapsulation in liposomes, microparticles, microcapsules, recombinant cells capable of expressing the therapeutic molecule(s) (see, *e.g.*, Wu *et al.*, *J. Biol. Chem.* 262, 4429, 1987), construction of a therapeutic nucleic acid as part of a retroviral or other vector, and the like. Methods of introduction include, but are not limited to, intrathecal, intradermal, intramuscular, intraperitoneal (ip), intravenous (iv), subcutaneous, intranasal, epidural, and oral routes. The therapeutics may be administered by any convenient route, including, for example, infusion or bolus injection, topical, absorption through epithelial or mucocutaneous linings (*e.g.*, oral mucosa, rectal and intestinal mucosa, and the like) ophthalmic, nasal, and transdermal, and may be administered together with other biologically active agents. Administration can be systemic or local. In addition, it may be desirable to introduce the pharmaceutical compositions of the invention by any suitable route, including intraventricular and intrathecal injection; intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir. Pulmonary administration can also be employed (*e.g.*, by an inhaler or nebulizer), for instance using a formulation containing an aerosolizing agent.

In a specific embodiment, it may be desirable to administer the pharmaceutical compositions of the invention locally to the area in need of treatment. This may be achieved by, for example, and not by way of limitation, local infusion or perfusion during surgery, topical application (*e.g.*, wound dressing), injection, catheter, suppository, or implant (*e.g.*, implants formed from porous, non-porous, or gelatinous materials, including membranes, such as sialastic membranes or fibers), and the like. In one embodiment, administration can be by direct injection at the site (or former site) of a tissue that is

to be treated. In another embodiment, the therapeutic are delivered in a vesicle, in particular liposomes (see, *e.g.*, Langer, *Science* 249, 1527, 1990; Treat *et al.*, in *Liposomes in the Therapy of Infectious Disease and Cancer*, Lopez-Berestein and Fidler (eds.), Liss, N.Y., pp. 353-365, 1989).

In yet another embodiment, the therapeutic can be delivered in a controlled release system. In one embodiment, a pump may be used (see, *e.g.*, Langer *Science* 249, 1527, 1990; Sefton *Crit. Rev. Biomed. Eng.* 14, 201, 1987; Buchwald *et al.*, *Surgery* 88, 507, 1980; Saudek *et al.*, *N. Engl. J. Med.* 321, 574, 1989). In another embodiment, polymeric materials can be used (see, *e.g.*, Ranger *et al.*, *Macromol. Sci. Rev. Macromol. Chem.* 23, 61, 1983; Levy *et al.*, *Science* 228, 190, 1985; During *et al.*, *Ann. Neurol.* 25, 351, 1989; Howard *et al.*, *J. Neurosurg.* 71, 105, 1989). Other controlled release systems, such as those discussed in the review by Langer (*Science* 249, 1527 1990), can also be used.

The vehicle in which the agent is delivered can include pharmaceutically acceptable compositions of the compounds, using methods well known to those with skill in the art. For instance, in some embodiments, small dsRNAs typically are contained in a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable" means approved by a regulatory agency of the federal or a state government or listed in the U.S. Pharmacopoeia or other generally recognized pharmacopoeia for use in animals, and, more particularly, in humans. The term "carrier" refers to a diluent, adjuvant, excipient, or vehicle with which the therapeutic is administered. Such pharmaceutical carriers can be sterile liquids, such as water and oils, including those of petroleum, animal, vegetable, or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil, and the like. Water is a preferred carrier when the pharmaceutical composition is administered intravenously. Saline solutions, blood plasma medium, aqueous dextrose, and glycerol solutions can also be employed as liquid carriers, particularly for injectable solutions. The medium may also contain conventional pharmaceutical adjunct materials such as, for example, pharmaceutically acceptable salts to adjust the osmotic pressure, lipid carriers such as cyclodextrins, proteins such as serum albumin, hydrophilic agents such as methyl cellulose, detergents, buffers, preservatives and the like.

Examples of pharmaceutical excipients include starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol, and the like. The therapeutic, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. These therapeutics can take the form of solutions, suspensions, emulsion, tablets, pills, capsules, powders, sustained-release formulations, and the like. The therapeutic can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. A more complete explanation of parenteral pharmaceutical carriers can be found in Remington: *The Science and Practice of Pharmacy* (19th Edition, 1995) in chapter 95.

Embodiments of other pharmaceutical compositions are prepared with conventional pharmaceutically acceptable counter-ions, as would be known to those of skill in the art.

Therapeutic preparations will contain a therapeutically effective amount of at least one active ingredient, preferably in purified form, together with a suitable amount of carrier so as to provide proper administration to the patient. The formulation should suit the mode of administration.

The composition of this disclosure can be formulated in accordance with routine procedures as a pharmaceutical composition adapted for intravenous administration to human beings. Typically, compositions for intravenous administration are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may also include a solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the injection.

The ingredients in various embodiments are supplied either separately or mixed together in unit dosage form, for example, in solid, semi-solid and liquid dosage forms such as tablets, pills, powders, liquid solutions, or suspensions, or as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the composition is administered by injection, an ampoule of sterile water or saline can be provided so that the ingredients may be mixed prior to administration.

The amount of the therapeutic that will be effective depends on the nature of the disorder or condition to be treated, as well as the stage of the disorder or condition. Effective amounts can be determined by standard clinical techniques. The precise dose to be employed in the formulation will also depend on the route of administration, and should be decided according to the judgment of the health care practitioner and each patient's circumstances. An example of such a dosage range is 0.1 to 200 mg/kg body weight in single or divided doses. Another example of a dosage range is 1.0 to 100 mg/kg body weight in single or divided doses.

The specific dose level and frequency of dosage for any particular subject may be varied and will depend upon a variety of factors, including the activity of the specific compound, the metabolic stability and length of action of that compound, the age, body weight, general health, sex, diet, mode and time of administration, rate of excretion, drug combination, and severity of the condition of the host undergoing therapy.

The dsRNAs of the present disclosure can be administered at about the same dose throughout a treatment period, in an escalating dose regimen, or in a loading-dose regime (*e.g.*, in which the loading dose is about two to five times the maintenance dose). In some embodiments, the dose is varied during the course of a treatment based on the condition of the subject being treated, the severity of the disease or condition, the apparent response to the therapy, and/or other factors as judged by one of ordinary skill in the art. In some embodiments long-term treatment with the drug is contemplated, for instance in order to reduce the occurrence of expression or overexpression of the target gene (*e.g.*, ErbB2).

In some embodiments, sustained intra-tumoral (or near-tumoral) release of the pharmaceutical preparation that comprises a therapeutically effective amount of dsRNA may be beneficial. Slow-release formulations are known to those of ordinary skill in the art. By way of

example, polymers such as bis(p-carboxyphenoxy)propane-sebacic-acid or lecithin suspensions may be used to provide sustained intra-tumoral release.

It is specifically contemplated in some embodiments that delivery is via an injected and/or implanted drug depot, for instance comprising multi-vesicular liposomes such as in DepoFoam (SkyePharma, Inc, San Diego, CA) (see, for instance, Chamberlain *et al.*, *Arch. Neuro.* 50:261-264, 1993; Katri *et al.*, *J. Pharm. Sci.* 87:1341-1346, 1998; Ye *et al.*, *J. Control Release* 64:155-166, 2000; and Howell, *Cancer J.* 7:219-227, 2001).

In other embodiments, perfusion of a tumor with a pharmaceutical composition that contains a therapeutically effective amount of an sRNA is contemplated, for instance an amount sufficient to provide a measurable reduction in tumor growth, tumor size, tumor cell growth, or another measurable reduction in the disease being treated.

IX. Combination Therapy

The present disclosure also contemplates combinations of dsRNAs with one or more other agents useful in the treatment of a disease, such as a hyper-proliferative disease. For example, dsRNAs may be administered in combination with effective doses of other medicinal and pharmaceutical agents. In some embodiments, one or more known anti-cancer drugs are included with a dsRNA that targets a gene known to be involved in a hyper-proliferative disorder, such as Erb-B2. The term "administration in combination with" refers to both concurrent and sequential administration of the active agents.

In addition, dsRNAs may be administered in combination with effective doses of radiation, anti-proliferative agents, anti-cancer agents, immunomodulators, anti-inflammatories, anti-infectives, hypomethylation agents, nucleosides and analogs thereof, and/or vaccines.

Examples of anti-proliferative agents that can be used in combination with a dsRNA (such as an dsRNA specific for ErbB2) include, but are not limited to, the following: ifosamide, cisplatin, methotrexate, procarizine, etoposide, BCNU, vincristine, vinblastine, cyclophosphamide, gencitabine, 5-fluorouracil, paclitaxel, or doxorubicin.

Non-limiting examples of immuno-modulators that can be used in combination with a dsRNA (such as an dsRNA specific for ErbB2) are AS-101 (Wyeth-Ayerst Labs.), bropirimine (Upjohn), gamma interferon (Genentech), GM-CSF (granulocyte macrophage colony stimulating factor; Genetics Institute), IL-2 (Cetus or Hoffman-LaRoche), human immune globulin (Cutter Biological), IMREG (from Imreg of New Orleans, La.), SK&F 106528, and TNF (tumor necrosis factor; Genentech).

Specific examples of specific compounds that in some embodiments are used in combination with a dsRNA (such as an dsRNA specific for ErbB2) are 5-azacytidine, 2'-deoxy-4-azacytidine, ara-C, and tricostatin A.

The combination therapies are of course not limited to the lists provided in these examples, but includes any composition for the treatment of diseases or conditions to which the dsRNA is targeted.

X. Gene Silencing Array Kits and Gene Silencing Kits

Gene silencing arrays as disclosed herein can be supplied in the form of a kit for use in gene function analyses. In such a kit, at least one gene-silencing array is provided. The kit also includes instructions, usually written instructions, to assist the user in using the array. Such instructions can optionally be provided on a computer readable medium.

Kits may additionally include one or more buffers or media for use during assay of the provided array. These solutions may be provided in bulk, where each container of buffer is large enough to hold sufficient buffer for several probing or washing or stripping procedures. Alternatively, the solutions can be provided in pre-measured aliquots, which would be tailored to the size and style of array included in the kit.

Certain kits provide one or more containers in which to carry out one or more of the reactions involved in culturing or analyzing the array.

Certain kits provide one or more live cultures of cells for transfection on the provided array.

Kits may in addition include either labeled or unlabeled control probe or transfection molecules, to provide for internal tests of either the transfection procedure or probing of the gene silencing array, or both. The control molecule(s) may be provided suspended in an aqueous solution or as a freeze-dried or lyophilized powder, for instance. The container(s) in which the control(s) are supplied can be any conventional container that is capable of holding the supplied form, for instance, microfuge tubes, ampoules, or bottles. In some applications, control probes may be provided in pre-measured single use amounts in individual, typically disposable, tubes or equivalent containers. Optionally, control molecules, particularly controls for transfection, can be provided printed onto the included array.

The amount of each component supplied in the kit can be any particular amount, depending for instance on the market to which the product is directed. For instance, if the kit is adapted for research or clinical use, sufficient component(s) likely will be provided to perform several transfections or analyses of the array(s).

In certain embodiments, a plurality of different control molecules, for instance for use in different endpoint analyses, will be provided in a single kit.

Gene silencing agents as disclosed herein also can be supplied in the form of a kit, for instance for use in altering gene expression. In such a kit, at least one gene-silencing agent, *e.g.*, a small double stranded RNA, is provided. The kit also includes instructions, usually written instructions, to assist the user in using the agent. Such instructions can optionally be provided on a computer readable medium. The agent in many embodiments is provided in a pharmaceutically acceptable carrier, such as those described herein.

Kits may additionally include one or more buffers or media for use during assay of the provided agent.

The amount of each component supplied in the kit can be any particular amount, depending for instance on the market to which the product is directed. For instance, if the kit is adapted for research or clinical use, sufficient component(s) likely will be provided to perform several treatments.

The following examples describe and illustrate certain particular features and/or embodiments. These examples are intended to be merely illustrative and not limiting in either scope or spirit. Unless indicated otherwise, all percentages are by weight. Those skilled in the art will readily understand that variations of the materials, conditions, and processes described in the examples can be used.

EXAMPLES

Example 1: Identification of small dsRNAs that mediate RNA interference (RNAi) in invertebrate and vertebrate systems

This example provides methods for the construction and analysis of short double-stranded RNAs in *C. elegans*, Drosophila, murine and human cells, as well as characterization of several different forms of small double-stranded RNAs.

Methods

Nucleic acids. Single stranded, gene specific sense and antisense RNA oligomers were synthesized using 2'-O-(tri-isopropyl) silyloxymethyl chemistry by Xeragon AG (Zurich, Switzerland). It has previously been shown that RNAs produced by this methodology are highly pure and efficiently form RNA duplexes (Parrish *et al.*, *Mol. Cell* 6, 1077, 2000; Caplen *et al.*, *Gene* 252, 95, 2000).

For studies conducted in *C. elegans*, RNA oligomers were annealed and injected into adults at a concentration of 5 mg/ml as described (Parrish *et al.*, *Mol. Cell* 6, 1077, 2000). For experiments conducted using mammalian cells, dsRNA molecules were generated by mixing sense and antisense ssRNA oligomers (100 g each) in 10 mM Tris-Cl (pH 7.0), 20 mM NaCl (total volume 300 μl), heating to 95° C, and cooling slowly (18 hours) to room temperature. The dsRNAs were ethanol precipitated and re-suspended in water at approximately 0.5 mg/ml. The integrity and the dsRNA character of the annealed RNAs were confirmed by gel electrophoresis. The sequences of the RNA oligonucleotides used in Example 1 are shown in Table 1; the *cat* 22 and 23 ssRNA oligomers were HPLC purified. Plasmid pEGFP-N3 (Clontech, Palo Alto, CA) expresses a mammalian enhanced version of green fluorescent protein (GFP) and neomycin phosphotransferase (neo). Plasmid pEGFPd2 (Clontech, Palo Alto, CA) expresses a mammalian enhanced version of green fluorescent protein (GFP) with a protein half-life of approximately 2 hours and neomycin phosphotransferase (neo). Plasmid pcDNA3.CAT (Invitrogen, Carlsbard, CA) expresses chloramphenicol acetyl transferase (CAT) and neo. Plasmids pAct.GFP and pAct.CAT are *Drosophila* expression vectors that express GFP and CAT respectively.

Cell culture and nucleic acid transfections. All mammalian cells were grown in Dulbeccos Modified Eagles Medium (DMEM, Life Technologies, Gaithersburg, MD) supplemented with 10% FBS (Gemini BioProducts Inc., Calabasas, CA). Primary mouse embryonic fibroblasts (MEFs) from wild-type I129 mouse embryos, (a gift of Dr. J. Bell, University of Ottawa, Ontario, Canada) (Abraham *et al.*, *J. Biol. Chem.* 274, 5953, 1999) were expanded to generate a more homogenous cell line and were used at passages 20 through 50 (Todaro *et al.*, *J. Cell Biol.* 17, 299, 1963). Cell line 293 is a human embryonic kidney cell line (Graham *et al.*, *J. Gen. Virol.* 36, 59, 1977); HeLa is a human epithelial cell line derived from a cervical adenocarcinoma (American Tissue Culture Collection (ATTC), Manassas, VA, ATTC#CCL-2).

Plasmid/RNA co-transfection of mammalian cells was mediated using the cationic lipid Lipofectamine and the propriety plus reagent (Life Technologies). Cells were seeded about 18 hours prior to transfection and were transfected at about 70-80% confluency. Plasmid DNA was complexed with the plus reagent (4-6 μl per 2 μg DNA) in DMEM for about 15 minutes. Five to ten minutes into the plasmid/plus reagent incubation, RNAs were added. Lipofectamine diluted in DMEM was added to the plasmid/plus reagent/RNA mixture and complexation was continued for an additional 15 minutes. The amount of Lipofectamine added (8-15 μg) was based on the total weight of nucleic acid (DNA and RNA) used and a weight-to-weight ratio of nucleic acid to lipid of 1:4. The amount of RNA used was adjusted to account for the variations in the sizes of RNA. For small RNAs (21-27 nt) 70 pmoles of ssRNA and dsRNA was used, corresponding to approximately 0.5 μg of a 22 nt ssRNA and 1 μg of 22 nt dsRNA. For the larger RNAs (78-81 nt) approximately 30 pmoles of RNA was used (0.85 μg of ssRNA and 1.7 μg of dsRNA). Three hours after initiation of transfection, DMEM supplemented with 20% FBS was added to cells.

S2 cells were grown in DES™ Medium (Invitrogen, Carlsbad, CA) supplemented with 10% FBS (Gemini BioProducts Inc., Calabasas, CA). Cells were passaged every two to three days to maintain exponential growth. S2 Cells were transfected using either the cationic lipid CellFectin (Life Technologies, Gaithersburg, MD), or DOTAP (Roche Biochemicals, Indianapolis, IN) using an adaptation of the manufacturer's protocol. Briefly, cells were seeded and allowed to settle overnight, nucleic acid (plasmid and/or dsRNA) was complexed with lipid at a weight to weight ratio of 1:6 in DES medium without supplementation. The complex was incubated at room temperature for 15 minutes and then added to cells from which normal growth medium had been removed. After overnight incubation, an equal volume of DES medium plus 20% FBS was added to the cell / lipoplex mixture.

Analysis of gene expression. The *C. elegans unc-22* gene encodes an abundant striated muscle component, loss of the expression (or function) of which results in a characteristic twitching phenotype. Animals were scored for the twitching phenotype as described in Parrish *et al.* (*Mol. Cell* 6, 1077, 2000). GFP expression was assessed in mammalian cells by fluorescence activated cell analysis (FACs) (FacsCaliber, Becton Dickinson, San Jose, CA) using pcDNA3.CAT transfected cells to control for background fluorescence. CAT expression was assessed using an ELISA-based

assay (Roche Biochemicals, Indianapolis, IN). Total protein was determined using the Bradford method as previously described (Caplen *et al.*, *Gene* 252, 95, 2000).

Poly A+ RNA was purified from MEFs using GTC extraction, Oligo dT cellulose chromatography, and DNase digestion to remove residual plasmid DNA. Following electrophoresis (1.2 % agarose, 1x MOPS, 5.0% formaldehyde) and Northern blot transfer, filters were sequentially hybridized with random prime labeled cDNA probes corresponding to *egfp* and *neo*. Hybridization intensities were measured using a BAS150 phosphoimager and pixel densities calculated using Image reader 1.4 and ImageGauge 3.0 (Fuji Ltd., Tokyo, Japan).

Cell survival and *in vitro* kinase assays. To assay cell survival, MEFs were plated in 96 well plates about 18 hours prior to transfection and were transfected at about 70-80% confluency using Lipofactamine as a carrier. RNA transfections were conducted as above, except for the omission of the plus complexation step, and using $1/10^{th}$ the amount of RNA and lipid and $1/10^{th}$ the volume of medium. Cell viability was determined 48 hours after initiation of transfection using the MTT labeling reagent as described by the manufacturer (Roche Biochemicals).

In vitro kinase reactions were conducted in a final volume of 12.5 ml using 100 mM [γ^{-32}P]-ATP (specific activity 1 Ci/mM; Amersham, Piscataway, NJ), 100 mM ATP (Sigma, St. Louis, MO) in 20 mM Hepes, pH 7.5, 90 mM KCl, 5 mM MgOAc, 1 mM dithiotreitol (DTT) and an equal amount of cell lysate prepared from 1 x 10^6 human Jurkat T lymphocytes treated with 100 u/ml of rhIFN-ß for 24 hours prior to lysis (lysis buffer: 20 mM Hepes, 120 mM KCl, 5 mM MgOAc, 1 mM Benzamidine, 1 mM DTT, 1% nonidet P-40). Double stranded RNA (1 µg/ml) was added to each reaction mixture and the reactions incubated for 10 minutes at 30° C. Reactions were quenched by addition of an equal volume of 2 X sample buffer (2 X sample buffer: 62.5 mM Tris-Cl pH 6.8, 10% glycerol, 2% sodium dodecyl sulfate, 0.0125% bromophenol blue, 5% ß-mercaptoethanol, boiled for 2 minutes and subjected to electrophoresis (10% SDS-PAGE). Labeled proteins were visualized by autoradiography of dried gels.

A series of dsRNAs were generated from chemically synthesized ssRNAs. The small dsRNAs varied from 20 to 27 nts and had sequences that matched four different target RNAs, *unc-22*, *cat*, *lacZ*, and *egfp*. The sequences for these RNA oligonucleotides, as well as several longer RNAs for comparison purposes, are included in Table 1.

Table 1: Sequences of *gfp*, *cat*, *unc22* and *LacZ* dsRNAs (used in Example 1)

SEQ ID No.	Name	Sequence	Location
1	*egfp* 21 nt sense	5' PO_4 r(gacguaaacggccacaaguuc) 3' OH	nt 64 - nt 84
2	*egfp* 21 nt antisense	5' PO_4 r(acuuguggccguuuacgucgc) 3' OH	nt 62 - nt 82
3	*egfp* 22 nt sense	5' PO_4 r(gcaagcugacccugaaguucau) 3' OH	nt 122 - nt 143
4	*egfp* 22 nt antisense	5' PO_4 r(gaacuucagggucagcuugccg) 3' OH	nt 120 - nt 141
5	*egfp* 23 nt sense	5' PO_4 r(agcagcacgacuucuucaagucc) 3' OH	nt 239 - nt 261
6	*egfp* 23 nt antisense	5' PO_4 r(acuugaagaagucgugcugcuuc) 3' OH	nt 237 - nt 259
7	*egfp* 24 nt sense	5' PO_4 r(caucuucuucaaggacgacggcaa) 3' OH	nt 294 - nt 317
8	*egfp* 24 nt antisense	5' PO_4 r(gccgucguccuugaagaagauggu) 3' OH	nt 292 - nt 315
9	*egfp* 25 nt sense	5' PO_4 r(gcacaagcuggaguacaacuacaac) 3' OH	nt 417 - nt 441
10	*egfp* 25 nt antisense	5' PO_4 r(uguaguuguacuccagcuucugccc) 3' OH	nt 415 - nt 439

11	*egfp* 26 nt sense	5' PO_4r(cgacaagcagaagaacggcaucaagg) 3' OH	nt 465 - nt 490
12	*egfp* 26 nt antisense	5' PO_4r(uugaugccguucuucugcuugucggc) 3' OH	nt 463 - nt 488
13	*egfp* 27 nt sense	5' PO_4r(cuucaagauccgccacaacaucgagga) 3' OH	nt 495 - nt 521
14	*egfp* 27 nt antisense	5' PO_4r(cucgauguuguggcggaucuugaaguu) 3' OH	nt 493 - nt 519
15	*cat* 22 nt sense	5' PO_4r (gagugaauaccacgacgauuuc) 3' OH	nt 318 - nt 339
16	*cat* 22 nt antisense	5' PO_4r(aaucgucgugguauucacucca) 3' OH	nt 316 - nt 337
17	*cat* 23 nt sense	5' PO_4r(ggagugaauaccacgacgauuuc) 3' OH	nt 317 - nt 339
18	*cat* 23 nt antisense	5' PO_4r(aaucgucgugguauucacuccag) 3' OH	nt 315 - nt 337
19	*unc22* 23 nt sense	5' PO_4r(ucugucucugcuccucucggcgg) 3' OH	nt 17893 - nt 17915
20	*unc22* 23 nt antisense	5' PO_4r(gccgagaggagcagagacagagg) 3' OH	nt 17891 - nt 17913
21	*unc22* 24 nt sense	5' PO_4r(cucugucucugcuccucucggcgg) 3' OH	nt 17892 - nt 17915
22	*unc22* 24 nt antisense	5' PO_4r(gccgagaggagcagagacagagga) 3' OH	nt 17890 - nt 17913
23	*unc22* 25 nt sense	5' PO_4r(ccucugucucugcuccucucggcgg) 3' OH	nt 17891 - nt 17915
24	*unc22* 25 nt antisense	5' PO_4r(gccgagaggagcagagacagaggau) 3' OH	nt 17889 - nt 1791
25	*Zr1 unc 22* dsRNA 26 nts sense	5' PO_4r(uccucugucucugcuccucucggcgg) 3' OH	nt 17890 - nt 17915
26	*Zr9 unc 22* dsRNA 27 nts sense	5' PO_4r(aacccaucuccaccaaauaacuacaau) 3' OH	nt 14287 – nt 14313
27	*Zr11 unc 22* dsRNA · 32 nts) sense	5' PO_4r(aaaaagcaagccaacgaaaccaaagggaccaa) 3' OH	nt 4974 – nt 5005
28	*Zr7 unc 22* dsRNA 37 nts sense	5' PO_4 r(aaguugaagguuaagaaugaguügggagaggaugaag) 3'	nt 9331 – nr 9367
29	*LacZ* 22 nt sense	5' PO_4r(gguggcgcuggaugguaagccg) 3' OH	nt 1961 – nt 1982
30	*LacZ* 22 nt antisense	5' PO_4r(gcuuaccauccagcgccaccau) 3' OH	nt 1959 - nt 1980
31	*LacZ* 23 nt sense	5' PO_4r(gguggcgcuggaugguaagccgc) 3' OH	nt 1961 - nt 1983
32	*LacZ* 23 nt antisense	5' PO_4r(ggcuuaccauccagcgccaccau) 3' OH	nt 1959 – nt 1981
39	*egfp* 78 nt sense	5' PO_4r(cuacguccaggagcgcaccaucuucuucaaggacgacggcaacuacaagacccgcgccgaggugaaguucgagggcga) 3' OH	nt 276-nt 354
40	*egfp* 78 nt antisense	5' PO_4r(ucgcccucgaacuucaccucggcgcgggucuuguaguugccgucguccuugaagaagauggugcgcuccuggacguag) 3' OH	nt 276- nt 354
41	*cat* 78 nt sense	5' r(caucgcucuggagugaauaccacgacgauuuccggcaguuucuacacauauauucgcaagaugugcguguuacggug) 3'	nt 313 – nt 392
42	*cat* 78 nt antisense	5' r(caccguaacacgccacaucuugcgaauauauguguagaaacugccggaaaucgucgugguauucacuccagagcgaug) 3'	nt 313 – nt 392
43	*unc22* 81 nt sense	5' r(gugcucggaaaaccaucuagcccauugggaccuuuggaagugucgaaugucuacgaagaucgcgcagauuuggaguggaaa) 3'	nt 4096 – nt 4176
44	*unc22* 85 nt antisense	5' r(guacuuuccacuccaaaucugcgcgaucuucguagacaaucgacacuuccaaaggucccaaugggcuagaugguuuuccgagcac) 3'	nt 4096 – nt 4181
45	*LacZ* 81 nt sense	5' r(agcuccugcacuggauggugcgcuggaugguaagccgcuggcaagcggugaagugccucuggaugucgcuccacaaggua) 3'	nt 1946 – nt 2027
46	*LacZ* 81 nt antisense	5' r(uaccuuguggagcgacauccagaggcacuucaccgcuugccagcggcuuaccauccagcgccaccauccagugcaggagcu) 3'	nt 1946 – nt 2027

Sequence location or positions in Table 3 refer to the 720 nt coding region of *egfp* gene, the 675 nt coding region of the *cat* gene, the 3066 nt coding region of the *β-galactosidase* gene, and 18147 nt coding region of the *C. elegans unc22* gene.

Results and Analysis

Short RNase III-like products can induce inhibition of gene expression in *C. elegans.*

To determine whether small dsRNAs can be used directly to inhibit gene expression we assessed interference in *C. elegans* using small dsRNAs corresponding to *C. elegans unc-22*. dsRNA

molecules were formed with each strand carrying a 5'-PO_4, 3'-OH, and 2 base 3' overhangs. These were injected into adult *C. elegans*. The results are shown in Table 2.

Table 2.

Injection	Fraction affected (number scored)
unc-22 dsRNA 23 nts	1.4% (145)
unc-22 dsRNA 24 nts	3.6% (279)
unc-22 dsRNA 25 nts	16.3% (768)
unc-22 sense ssRNA 25 nts	0% (>1100)
unc-22 antisense ssRNA 25 nts	0% (>600)
unc-22 dsRNA 81 nts	88.9% (180)
egfp dsRNA 22 nts	0% (>300)
egfp dsRNA 23 nts	0% (>300)
egfp dsRNA 24 nts	0% (>300)
egfp dsRNA 25 nts	0% (>300)
No Injection	0% (>300)
Zr1 unc-22 dsRNA 26 nts (no adenosine residues)	21% (124)
Zr9 unc-22 dsRNA 27 nts (no guanosine residues)	0% (600)
Zr11 unc-22 dsRNA 32 nts (no uracil residues)	9.2% (174)
Zr7 unc-22 dsRNA 37 nts (no cytosine residues)	32.0% (100)

Percentages shown denote portion of progeny broods that show a specific decrease in *unc-22* function as evidenced by twitching behavior in 330 μM levamisole. Numbers in parenthesis are total numbers of animals scored.

unc-22 provides a sensitive and specific assay for genetic interference, as this is the only gene in the *C. elegans* genome that can mutate by loss of function to give a twitching phenotype. *unc-22* small dsRNAs induced a decrease in *unc-22* gene expression as measured by the presence of the twitching phenotype in the progeny of injected adults. Small dsRNAs of 23, 24, 25, 26, 32, and 37 nts produced interference. The 27 nt small dsRNA lacks guanosine residues and this may be the reason this small dsRNA failed to mediate interference. As a control, dsRNAs directed against an unrelated sequence (*egfp*) induced no phenotypic changes.

The effect of small double stranded RNAs on gene expression in *Drosophila* embryonic S2 cells.

The effect of 21-27 nt *egfp* dsRNAs on GFP expression following co-transfection of RNAs and the *Drosophila* plasmid pAct.GFP was assessed in *Drosophila* embryonic S2 cells (Fig. 1). Consistent with our previous observations of RNAi in S2 cells the *egfp* 78 nt dsRNA significantly inhibited GFP expression (by approximately 85%) (Caplen *et al.*, *Gene* 252, 95, 2000). The *egfp* dsRNAs of 21-24 nts had minimal effect on GFP expression, but the *egfp* dsRNAs of 25-27 nts reduced the percentage of cells expressing GFP by approximately 20% and the intensity of GFP fluorescence intensity by approximately 40%. The single stranded sense and antisense *egfp* ssRNAs had no significant effect on GFP expression.

21-23 nt dsRNAs inhibit expression in mouse embryonic fibroblasts.

To test if small dsRNA molecules can specifically inhibit gene expression in vertebrate cells mouse embryonic fibroblasts were co-transfected with expression plasmids encoding GFP (pEGFP-N3) and CAT (pcDNA3.CAT) and synthetic small dsRNAs corresponding to *egfp*, *cat*, or *unc-22*

(Fig. 2). The *egfp* dsRNAs (21-27 nts) all inhibited GFP expression in MEFs. The 22 and 23 nt *egfp* dsRNAs (20 and 21 nts base-paired with 2 nt 3' overhangs) showed the greatest degree of inhibition, both with respect to the total number of cells expressing GFP (Fig. 2A) and the fluorescence intensity of the GFP expression observed in GFP positive cells (Fig. 2B). In contrast, *unc-22* dsRNAs of 23 – 25 nts had no significant effect on GFP expression (Fig. 2C and D).

To further assess the efficacy and specificity of the inhibition mediated by dsRNAs in mammalian cells we used a second reporter, chloramphenicol acetyl transferase (Fig. 2E-I). *cat* dsRNAs of 22 and 23 nts completely inhibited CAT expression (Fig. 2E and F) while *unc-22* and *egfp* dsRNAs had no little or no effect on CAT expression (Fig. 2G-I). Although no antisense effect had been seen using GFP as a reporter, the *cat* ssRNA antisense oligomers partially inhibited CAT expression. However, the small dsRNA-mediated inhibition was more potent (about 1.5 fold) suggesting that the gene silencing mediated by the small dsRNAs can be distinguished from a purely antisense-based mechanism.

To analyze this inhibition of *egfp* expression at an RNA level, poly A+ RNA was purified from transfected MEFs and subjected to Northern analysis using cDNA probes corresponding to *egfp* and *neo,* both encoded by the pEGFP-N3 plasmid (Fig. 2J). Quantitative phosphoimage analysis showed a decrease in the levels of the *egfp* mRNA obtained from cells co-transfected with the pEGFP-N3 plasmid and the 21, 22, and 23 nt *egfp* small dsRNAs compared to cells transfected with the GFP plasmid alone. The percentage decrease was approximately 60% for all three *egfp* small dsRNAs when compared to the levels of *egfp* mRNA in cells transfected with plasmid only. Importantly, no effect was seen on the levels of the *neo* transcript compared to plasmid only transfected cells indicating that the inhibition induced by the small *egfp* dsRNAs was sequence specific. Consistent with this hypothesis, the 23 nt dsRNA corresponding to the *C. elegans unc-22* gene had no effect on either *egfp* or *neo* expression.

To follow the fate of cells transfected with dsRNAs, MEF cell survival was assayed (Fig. 2K). Longer dsRNAs (78 or 81 nts with flush ends; see Table 3) induced a substantial degree of cell death (up to 50%) in a 48-hour period, whereas the small dsRNAs had a minimal effect on the growth of cells. Examining the effect of the larger dsRNAs on gene expression, it was observed that the larger dsRNAs (78 or 81 nts) induced a sequence non-specific decrease of 75% in the percentage of cells expressing GFP (Fig 2L) and in CAT protein levels, compared to plasmid controls. This non-specific decrease in expression is consistent with previous data from numerous mammalian systems and contrasts with the specific gene silencing the 78 nt *egfp* dsRNA induces in *Drosophila* S2 cells (Caplen *et al.*, *Gene* 252, 95, 2000).

The decrease in transgene expression following small dsRNA transfection could be distinguished from the non-specific inhibition by examining the GFP fluorescence intensity seen in viable cells. The fluorescence intensity of GFP expression best illustrates a change in the total amount of GFP made by a live cell and therefore is less influenced by non-specific cell death. Though some decrease (approximately 60%) in the fluorescence intensity was seen using the larger (about 80 nt) dsRNA molecules (irrespective of sequence), the *egfp* small dsRNAs of 22 and 23 nt

consistently reduced the intensity of the GFP signal (by approximately 90%) to near background levels (Fig. 2M). The difference in specificity between the longer dsRNAs and small dsRNAs could also be seen at an RNA level, where the 78 nt *egfp* dsRNA induced a significant decrease in both the *egfp* and *neo* transcripts, whereas the small dsRNAs only inhibited *egfp* (Fig. 2J).

Inhibition of gene expression in human somatic cells.

Prior to the work reported here, there had been no evidence of an RNAi-like process occurring in human somatic cells. To determine if small dsRNAs could also specifically inhibit gene expression in human cells, we co-transfected two commonly used human cell lines, the embryonic kidney cell line 293, and the epithelial carcinoma cell line HeLa with plasmids and RNA (Fig. 3). All of the *egfp* small dsRNAs tested inhibited GFP gene expression in 293 (Fig. 3A and B) and HeLa (Fig. 3C and D) cells, with the 22 and 23 nt *egfp* small dsRNAs inducing the greatest decrease in GFP expression. In 293 cells co-transfected with pEGFP-N3 and the 22 nt *egfp* small dsRNA the intensity of GFP expression was reduced to near background levels (Fig. 3B). Similar results were seen in HeLa cells co-transfected with pEGFP-N3 and the 22 or 23 nt *egfp* small dsRNAs (Fig. 3D). DsRNAs corresponding to *unc-22* had no effect on GFP expression in these cells.

The small dsRNA-triggered inhibition of GFP expression was dose dependent in that doubling the amount of dsRNA (from 70 to 140 pmoles) decreased GFP intensity by an additional 25% for the *egfp* 22 nt small dsRNA and by 45% for the *egfp* 23 nt small dsRNA.

CAT expression was also significantly inhibited by small dsRNAs corresponding to *cat* (Fig. 3E and F) in HeLa cells. Again, the inhibition mediated by the small dsRNAs was significantly higher than that seen using ssRNA antisense oligomers. Co-transfection of the pcDNA3.CAT plasmid and the *egfp* small dsRNAs of the same size and of similar GC/AT complexity had no effect on CAT expression (Fig. 3G and H).

Small dsRNA mediated inhibition of gene expression is independent of non-specific interference pathways activated by larger dsRNAs.

It has been previously reported that small blunt-ended dsRNAs of less than 30 bp do not activate PKR (Manche *et al.*, *Mol. Cell Biol.* 12, 5238, 1992). Indeed, at high concentrations these short dsRNAs can competitively inhibit activation of PKR by larger dsRNAs. Similarly, the synthetic dsRNAs used in this study did not activate PKR (Fig 4A), and inhibited the activation of PKR by a large viral dsRNA (Fig. 4B). Interestingly, in this assay activation of PKR by the 78 and 81 nt dsRNAs was undetectable, despite observing a substantial level of cell death. This suggests that other dsRNA-dependent kinases or other pathways may be contributing in MEFs to the decrease in gene expression and cell death observed with these RNAs.

To see if the small dsRNAs could block the toxic effect of the larger dsRNAs in cells, 293 cells were co-transfected with the pEGFP-N3 plasmid, the *egfp* 78 nt dsRNA, and either the *unc-22* 23 or 24 nt small dsRNAs or the *egfp* 23 or 24 nt small dsRNAs (Fig. 4C-F). The cell death induced by the 78 nt *egfp* dsRNA was not inhibited by the *unc-22* or *egfp* small dsRNAs (Fig. 4C and D versus Fig. 4E and F) but importantly the 78 nt *egfp* dsRNA did not block the specific inhibition of GFP expression mediated by the 23 nt *egfp* small dsRNA. Thus, it appears that the small dsRNA-

mediated gene silencing mechanism is independent of non-specific responses of mammalian cells to dsRNA.

Example 2: Characterization and optimization of small dsRNA mediate RNAi in invertebrate and vertebrate systems

As demonstrated herein, gene silencing in mammalian cells can be mediated by small double stranded RNAs (dsRNAs) of ~21-23 nucleotides (nts). Naturally occurring small interfering RNAs (siRNAs) have characteristic termini; however, little is known as to what features are critical for an exogenous small dsRNA that mediates RNAi in invertebrate and vertebrate cells. This Example explores variations in small dsRNA structure and how those variations alter RNA interference.

Variations in the degree of interference could be seen to be dependent on the length of the small dsRNA, the length and orientation of any terminal overhang and the sequence of the region of the mRNA targeted. Further, a variety of whole strand and terminal chemical modifications showed an asymmetric role for the two strands in RNAi as modification of the sense strand of the RNA duplex was more tolerated than alteration of the antisense strand. Using fluorophore tagged small dsRNAs a correlation between duplex uptake and silencing could be seen with phenotypic effects detected as little as six hours after transfection. In contrast to RNAi in lower eukaryotes, small dsRNA mediated gene silencing in mammalian cells was transient lasting only a few days; there was also no evidence of amplification or cell to cell spread of the gene silencing effect. These studies give insights into the process of RNAi in mammalian cells and assist in improving the design of small dsRNAs.

Methods

Nucleic acids. Single stranded, gene specific sense and antisense oligoribonucleotides were synthesized using 2'-O-(tri-isopropyl) silyloxymethyl chemistry by Xeragon Inc. (Huntsville, AL; Gaithersburg, MD). Double stranded RNA molecules were generated by mixing sense and antisense ssRNA oligomers (50-200 μg each) in 10 mM Tris-Cl (pH 7.0), 20 mM NaCl (total volume 300 μl), heating to 95°C, and by incubating for 1 hr at 37°C or by cooling slowly (18 hours) to room temperature. The integrity and the dsRNA character of the annealed RNAs were confirmed by gel electrophoresis. Plasmid pd2EGFP-N3 (Clontech, Palo Alto, CA) expresses a destabilized version of the mammalian enhanced version of green fluorescent protein (GFP) and neomycin phosphotransferase (neo). Sequences of small dsRNAs are shown in Table 3 and in the appropriate figures.

Table 3: Sequences for RNAs used to generate dsRNAs used in Example 2

SEQ ID NO	Name	Sequence	Location
1	*egfp* Region A 21 nt sense	5'PO_4 r(gacguaaacggccacaaguuc)3'OH	nt 64 - nt 84
2	*egfp* Region A 21 nt antisense	5'PO_4 r(acuuguggccguuuacgucgc)3'OH	nt 62 - nt 82
3	*egfp* Region B#1 22 nt sense	5'PO_4 r(gcaagcugacccugaaguucau)3'OH & 5'r(gcaagcugacccugaaguucau)3'OH	nt 122 - nt 143

4	*egfp* Region B#1 22 nt antisense	5'PO_4r(gaacuucagggucagcuugccg)3'OH & 5'r(gaacuucagggucagcuugccg)3'OH	nt 120 - nt 141
5	*egfp* Region C 23 nt sense	5'PO_4r(agcagcacgacuucuucaagucc)3'OH	nt 239 - nt 261
6	*egfp* Region C 23 nt antisense	5'PO_4r(acuugaagaagucgugcugcuuc)3'OH	nt 237 - nt 259
7	*egfp* Region D 24 nt sense	5'PO_4r(caucuucuucaaggacgacggcaa)3'OH	nt 294 - nt 317
8	*egfp* Region D 24 nt antisense	5'PO_4r(gccgucguccuugaagaagauggu)3'OH	nt 292 - nt 315
101	*egfp* Region D 23 nt sense	5' PO_4r(caucuucuucaaggacgacggca)3'OH	
102	*egfp* Region D 23 nt antisense	5' PO_4r(ccgucguccuugaagaagauggu)3'OH	
103	*egfp* Region D 22 nt sense	5' PO_4r(caucuucuucaaggacgacggc)3'OH	
104	*egfp* Region D 22 nt antisense	5' PO_4r(cgucguccuugaagaagauggu)3'OH	
105	*egfp* Region D 21 nt sense	5' PO_4r(caucuucuucaaggacgacgg)3'OH	
106	*egfp* Region D 21 nt antisense	5' PO_4r(gucguccuugaagaagauggu)3'OH	
37	*egfp* Region D 20 nt sense	5' PO_4r(caucuucuucaaggacgacg)3'OH	
38	*egfp* Region D 20 nt antisense	5' PO_4r(ucguccuugaagaagauggu)3'OH	
33	*egfp* Region B 22 nt 2 nt 5' overhang antisense	5' PO_4r(aguugaacuucagggucagcuu)3'OH	nt 124 - nt 145
34	*egfp* Region B 22 nt blunt termini antisense	5' PO_4r(augaacuucagggucagcuugc)3'OH	nt 122 - nt 143
35	*egfp* Region B 23 nt antisense	5' PO_4r(gaacuucagggucagcuugccgu)3'OH	nt 119 - nt 141
36	*egfp* Region B 21 nt antisense	5' PO_4r(aacuucagggucagcuugccg)3'OH	nt 120 - nt 140
107	*egfp* Region B #3	5'r(cggcaagcugacccugaaguuc)3'	
108	*egfp* Region B #4	5'r(ACUUCAGGGUCAGCUUGCCGUA)3'	
109	*egfp* Region B #5	5'r(aucugcaccaccggcaagcugc)3'	
110	*egfp* Region B #6	5'r(AGCUUGCCGGUGGUGCAGAUGA)3'	
99	*egfp* Region B #7	5'r(caucugcaccaccggcaagcug)3'	
100	*egfp* Region B #8	5'r(GCUUGCCGGUGGUGCAGAUGAA)3'	
97	*egfp* Region B #9	5'r(cccugaaguucaucugcaccac)3'	
98	*egfp* Region B #10	5'r(GGUGCAGAUGAACUUCAGGGUC)3'	
111	*egfp* Region B #11	5'r(caagcugacccugaaguucauc)3'	
112	*egfp* Region B #12	5'r(UGAACUUCAGGGUCAGCUUGCC)3'	
113	*egfp* Region B #13	5'r(aagcugacccugaaguucaucu)3'	
114	*egfp* Region B #14	5'r(AUGAACUUCAGGGUCAGCUUGC)3'	
115	*egfp* Region B #15	5'r(ucaucugcaccaccggcaagcu)3'	
116	*egfp* Region B #16	5'r(CUUGCCGGUGGUGCAGAUGAAC)3'	
117	*egfp* Region B #17	5'r(agcugacccugaaguucaucug)3'	
118	*egfp* Region B #18	5'r(GAUGAACUUCAGGGUCAGCUUG)3'	
119	*egfp* Region B #19	5'r(uacggcaagcugacccugaagu)3'	
120	*egfp* Region B #20	5'r(UUCAGGGUCAGCUUGCCGUAGG)3'	
121	*egfp* Region B #21	5'r(acggcaagcugacccugaaguu)3'	
122	*egfp* Region B #22	5'r(CUUCAGGGUCAGCUUGCCGUAG)3'	
123	*egfp* Region B #23	5'r(gcugacccugaaguucaucugc)3'	
124	*egfp* Region B #24	5'r(AGAUGAACUUCAGGGUCAGCUU)3'	
125	*egfp* Region B #25	5'r(cugacccugaaguucaucugca)3'	
126	*egfp* Region B #26	5'r(CAGAUGAACUUCAGGGUCAGCU)3'	
127	*egfp* Region B #27	5'r(ggcaagcugacccugaaguuca)3'	
128	*egfp* Region B #28	5'r(AACUUCAGGGUCAGCUUGCCGU)3'	
129	*egfp* Region B #29	5'r(cugcaccaccggcaagcugccc)3'	
130	*egfp* Region B #30	5'r(GCAGCUUGCCGGUGGUGCAGAU)3'	
131	*egfp* Region B #31	5'r(acccugaaguucaucugcacca)3'	
132	*egfp* Region B #32	5'r(GUGCAGAUGAACUUCAGGGUCA)3'	
133	*egfp* Region B #1 Sense uu 3' overhang	5'PO_4r(gcaagcugacccugaaguucuu)3'OH	
134	*egfp* Region B #1 Antisense uu 3' overhang	5'PO_4r(gaacuucagggucagcuugcuu)3'OH	
135	*egfp* Region B #1 Sense dTdT 3' overhang	5'PO_4r(gcaagcugacccugaaguucdTdT)3'OH	
136	*egfp* Region B #1 Antisense dTdT 3' overhang	5'PO_4r(gaacuucagggucagcuugcudTdT)3'OH	
137	*egfp* Region B #1 Sense 2'O-me modification	5'$PO_4$2'O-me r(GCAAGCUGACCCUGAAGUUCAU) 3'	
138	*egfp* Region B #1 Antisense 2'O-me modification	5'$PO_4$2'O-me-r(GAACUUCAGGGUCAGCUUGCCG) 3'	
139	*egfp* Region B #1 Sense no 5'PO_4	5'r(gcaagcugacccugaaguucau)3'OH	

140	*egfp* Region B #1 Antisense no 5'PO_4	5'r(gaacuucagggucagcuugccg)3'OH	
141	*egfp* Region B #1 Sense 5'Rh 3' OH	5'Rhodamine-r(gcaagcugacccugaaguucau)	
142	*egfp* Region B #1 Antisense 5'Rh	5'Rhodamine-r(gaacuucagggucagcuugccg)3'OH	
143	*egfp* Region B #1 Sense 3'Rh	5'PO_4r(gcaagcugacccugaaguucau)3'Rhodamine	
144	*egfp* Region B #1 Antisense 3'Rh	5'PO_4r(gaacuucagggucagcuugccg)3'Rhodamine	
145	*egfp* Region B #1 Sense gfp mut	5'PO_4r(gcaagcugacccacuaguucau)3'OH	
146	*egfp* Region B #1 Antisense gfp mut	5'PO_4r(gaacuagugggucagcuugccg)3'OH	
15	*cat* 22 nt sense	5' PO_4r(gagugaauaccacgacgauuuc)3'OH	nt 318 - nt 339
147	*cat-FITC* 22 nt sense	5' PO_4r(gagugaauaccacgacgauuuc)3'FITC	nt 318 - nt 339
16	*cat* 22 nt antisense	5' PO_4r(aaucgucgugguauucacucca)3'OH	nt 316 - nt 337
17	*cat* 23 nt sense	5' PO_4r(ggagugaauaccacgacgauuuc)3'OH	nt 317 - nt 339
18	*cat* 23 nt antisense	5' PO_4r(aaucgucgugguauucacuccag)3'OH	nt 315 - nt 337
29	*LacZ* 22 nt sense	5' PO_4r(gguggcgcuggaugguaagccg)3'OH	nt 1961 – nt 1982
30	*LacZ* 22 nt antisense	5' PO_4r(gcuuaccauccagcgccaccau)3'OH	nt 1959 - nt 1980
31	*LacZ 23* nt sense	5' PO_4r(gguggcgcuggaugguaagccgc)3'OH	nt 1961 - nt 1983
32	*LacZ 23* nt antisense	5' PO_4r(ggcuuaccauccagcgccaccau)3'OH	nt 1959 – nt 1981
148	*LacZ* Sense Z1	5'PO_4r(acccuggcguuacccaacuuaa)3'OH	nt 126 – nt 147
149	*LacZ* Antisense Z1	5'PO_4r(aaguuggguaacgccaggguuu)3'OH	
150	*LacZ* Sense Z2	5'PO_4r(gcuggcuggagugcgaucuu)3'OH	nt 290 – nt 309
151	*LacZ* Antisense Z2	5'PO_4r(gaucgcacuccagccagcuu)3'OH	
152	*LacZ* Sense Z3	5'PO_4r(ccuaucccauuacggucaaucc)3'OH	nt 388 - nt 409
153	*LacZ* Antisense Z3	5'PO_4r(auugaccguaaugggauagg)3'OH	
154	*LacZ* Sense Z4	5'PO_4r(ccgacuacacaaaucagcgauu)3'OH	nt 731 - nt 752
155	*LacZ* Antisense Z4	5'PO_4r(ucgcugauuuguguagucgguu)3'OH	
156	*LacZ* Sense Z5	5'PO_4r(guucagaugugcggcgaguu)3'OH	nt 811 - nt 830
157	*LacZ* Antisense Z5	5'PO_4r(cucgccgcacaucugaacuu)3'OH	
158	*LacZ* Sense Z6	5'PO_4r(cuuuaacgccgugcgcuguu)3'OH	nt 1232 - nt 1251
159	*LacZ* Antisense Z6	5'PO_4r(cagcgcacggcguuaaaguu)3'OH	
160	*LacZ* Sense Z7	5'PO_4r(gccaauauugaaacccacgg)3'OH	nt 1322 - nt 1341
161	*LacZ* Antisense Z7	5'PO_4r(guggguuucaauauuggcuu)3'OH	
162	*LacZ* Sense Z8	5'PO_4r(cugugccgaaaugguccaucaa)3'OH	nt 1628 - nt 1649
163	*LacZ* Antisense Z8	5'PO_4r(gauggaccauuucggcacagcc)3'OH	
164	*LacZ* Sense Z9	5'PO_4r(gcaaaacaccagcagcaguu)3'OH	nt 1947 - nt 1966
165	*LacZ* Antisense Z9	5'PO_4r(cugcugcugguguuuugcuu)3'OH	
166	*LacZ* Sense Z10	5'PO_4r(gugaccagcgaauaccuguu)3'OH	nt 2003 - nt 2022
167	*LacZ* Antisense Z10	5'PO_4r(cagguauucgcuggucacuu)3'OH	

Cell culture, nucleic acid transfections and analysis of gene expression. HeLa cells (a human epithelial cell line derived from a cervical carcinoma), American tissue culture collection (ATTC), Manassas, VA, ATTC#CCL-2) were grown in Dulbeccos Modified Eagles Medium (DMEM, Life Technologies, Gaithersburg, MD) supplemented with 10% FBS (Gemini BioProducts Inc., Calabasas, CA) (D10)). The human embryonic kidney cell line 293 (Graham *et al.*, *J. Gen. Virol.* 36, 59, 1977) and the mouse fibroblast Nih-3t3 cell line were grown under similar conditions. 293/d2eGFP and HeLa/d2eGFP cells stably expressing the destabilized version of EGFP were generated by cationic lipid transfection of the pd2EGFP-N3 (Clontech Inc., Palo Alto, CA) plasmid and selection with 1 mg/ml G418.

To obtain a population of GFP positive cells with a narrow range of fluorescence intensity cells were subjected to flow cytometric sorting. HeLa/d2eGFP cells were transfected with small dsRNAs using the cationic lipid Lipofectin (Life Technologies, Gaithersburg, MD). Cells were

seeded ~18 hours prior to transfection and were transfected at ~70-80% confluency. RNA (0.5-2 μg) was complexed with Lipofectin at a nucleic acid: cationic lipid ratio of 1:2 in OptiMEM for ~15 minutes. Three to four hours after initiation of transfection, DMEM supplemented with 20% FBS was added to cells. These transfection conditions represent an optimized set of condition; several other cationic lipid formulas and ratios of small dsRNA to cationic lipid were also tested and showed successful transfection of small dsRNAs and subsequent RNAi. 293/d2eGFP cells were transfected as described for HeLa/d2eGFp cells but using the cationic lipid Lipofectin at a ratio of 1:4 RNA to lipid. Murine 3T3/ß-gal cells stably expressing the *E. coli* ß-galactosidase gene were generated by retroviral mediated gene transfer of the mouse fibroblast cell line, NIH-3T3, and selection with 1mg/ml G418. The transfection of 3T3/ß-gal cells with small dsRNAs was conducted as described for Hela/d2 egfp cells except that the cationic lipid Lipofectamine (Life Technologies) at a RNA;lipid ratio of 1:5 was used as the transfection agent. DNA/dsRNA transfections of mouse embryonic fibroblasts were conducted as described in Example 1.

Analysis of gene expression. GFP expression was assessed in mammalian cells by fluorescence activated cell analysis (FACs) (FacsCaliber, Becton Dickinson, San Jose, CA) using untransfected cells to control for background fluorescence. ß-galactosdiase activity was assessed using a luminescent ß-galactosidase Detection Kit (Clontech Inc., Palo Alto, CA). Fluorescent images of cells were taken using an AxioPlan2 microscope fitted with a Marchauser 8 position stage and an Axiocam CCD digital camera (Carl Zeiss GmbH, Oberkochen, Germany) at 25 fold magnification.

Results

Inhibition of gene expression by small dsRNAs in a transient expression system is predictive of the degree of silencing seen in a stable expression system.

The studies documenting a RNAi like response mediated by small dsRNAs of approximately 21 or 22 nts primarily used co-transfection of dsRNAs and plasmid DNA expressing a transgene. To establish that the some of the same small dsRNAs can inhibit stable gene expression we analyzed the effect of transfecting the four small dsRNAs (Fig. 5A; SEQ ID NOs: 1, 3, 5, and 7) into HeLa cells that stably express a destabilized version of eGFP with a two-hour half-life, HeLa/d2eGFP cells (Fig. 5B and 5C). In both the transient (Example 1) and stable expression systems (Fig. 5) the rank order for the effectiveness of the four small dsRNAs was similar with the Region B 22 nt dsRNA (SEQ ID NOs: 3 and 4) the most effective and the Region D 24 nt dsRNA (SEQ ID NOs: 7 and 8) the least. The 21 and 23 nt dsRNAs corresponding to region A (SEQ ID NOs: 1 and 2) and C (SEQ ID NOs: 5 and 6) of egfp respectively showing an intermediate effectiveness, in that a significant change in the fluorescence intensity of GFP was seen but the percentage of cells expressing no GFP was less than that seen with the 22 nt dsRNA against Region B.

The size and sequence of a small dsRNA influences its' ability to mediate gene silencing

Analysis of gene expression using flow cytometry for eGFP is a highly sensitive assay for RNAi, because it is possible to quantify both the percentage of cells expressing eGFP and to estimate

the fluorescence intensity of eGFP expression in the entire population of cells, which should reflect the total level of eGFP protein. To examine if the variations seen in the ability to inhibit *egfp* expression was related to the size of the RNA duplex and / or composition of the target sequence, we first synthesized small dsRNAs corresponding to Region D of the *egfp* mRNA (Fig. 5A) decreasing in size by one nucleotide from the 3' end of the original 24 nt sequence. Representative histograms (Fig. 6A) show that a shift in the peak GFP fluorescence is seen as the size of the dsRNA is reduced, however, it was not until a duplex of 20 nts was used that a substantial increase in the number of completely eGFP negative cells is seen. The difference between the effectiveness of the 20 versus 21 nt dsRNAs is of particular note as these differ by only one nucleotide on the 3' end of the sequence and yet show dramatically different abilities to mediate inhibition of eGFP (Fig. 6B).

To analyze the effects of the sequence of an individual small dsRNAs' ability to mediate RNAi we analyzed a series of small dsRNAs covering a 50 nucleotide region of *egfp* that included the "Region B" *egfp* sequence (Fig. 5A and 5C). The ability of 16 pairs of siRNAs (all with 2 nt 3' overhangs) to knockdown eGFP expression was analyzed 72 hours after transfection of Hela/d2eGFP cells (Table 4). Of the 16 small dsRNAs tested, six inhibited both the percentage of eGFP positive cells (over 65% inhibition) and the fluorescence intensity of eGFP expression (over 90% inhibition). Five small dsRNAs showed an intermediate effect on eGFP expression (20 to 60% inhibition of the percentage of eGFP positive cells, 30-90% inhibition of eGFP fluorescence intensity). Four small dsRNAs showed little or no ability to mediate RNAi against *egfp* (less than 20% inhibition of the percentage of eGFP positive cells, less than 30% inhibition of eGFP fluorescence intensity). With one exception (SEQ ID NOs: 110 and 111) the degree of RNAi appeared to correlate closely with the region of the *egfp* sequence targeted.

To determine what characteristics of these small dsRNA lead to this difference in the ability to mediate RNAi against *egfp*, we considered a number of factors including, the secondary structure of the regions of the *egfp* mRNA targeted by the different small dsRNAs, the putative cleavage site within each small dsRNA, and the nucleotide compositions of the different small dsRNAs.

The secondary structure of a region targeted by an antisense single stranded RNA (ssRNA) can significantly influence the degree of down regulation of gene expression observed (Lebedeva & Stein, *Ann. Rev. Pharma. Toxic.* 41:403-419, 2001; Agrawal & Kandimalla, *Mol. Medicine Today* 6:72-81, 2000). To assess if the secondary structure of the region targeted by a particular small dsRNAs correlated with the its' ability to mediate inhibition of gene expression we aligned the sequences of the dsRNAs shown in Table 2 with the two primary m-folds of the *egfp* transcript (available from the mfold server: © 1995-2002, Michael Zuker, Rensselaer Polytechnic Institute) (Fig. 6C). No obvious secondary structural feature of the *egfp* transcript could be seen that correlated with the degree of inhibition of eGFP expression mediated by the each of the dsRNAs. This

Table 4

SEQ ID NO.	Region B *egfp*	% GFP Positive cells	% inhibition of eGFP expression	eGFP fluorescence intensity	% inhibition of eGFP fluorescence intensity	%/ % AU/GC	20 nt dsRNA AU/GC
312	cctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgccc						
119 *120R*	uacggcaagcugacccugaagu GGAUGCCGUUCGACUGGGACUU	24.33±0.81	71.7	8.85±0.37	91.1	45.55/54.55 40.91/59.09	45.0/55.0
121 *122R*	acggcaagcugacccugaaguu GAUGCCGUUCGACUGGGACUUC	28.60±2.384	66.7	10.22±0.74	89.7	45.45/54.55 40.91/59.09	40.0/60.0
107 *108R*	cggcaagcugacccugaaguuc AUGCCGUUCGACUGGGACUUCA	84.31±3.11	1.8	67.69±9.01	31.7	40.91/59.09 45.45/54.55	40.0/60.0
127 *128R*	ggcaagcugacccugaaguuca UGCCGUUCGACUGGGACUUCAA	20.74±0.9	75.8	8.25±0.46	91.7	45.45/54.55 45.45/54.55	45.0/55.0
3 *4R*	gcaagcugacccugaaguucau GCCGUUCGACUGGGACUUCAAG	26.71±0.64	68.9	9.75±0.14	90.2	50.00/50.00 40.91/59.09	45.0/55.0
111 *112R*	caagcugacccugaaguucauc CCGUUCGACUGGGACUUCAAGU	14.31±3.01	83.3	6.23±0.75	93.7	50.00/50.00 45.45/54.55	50.0/50.0
113 *114R*	aagcugacccugaaguucaucu CGUUCGACUGGGACUUCAAGUA	14.71±0.40	82.9	6.21±0.12	93.7	54.55/45.45 50.00/50.00	55.0/45.0
117 *118R*	agcugacccugaaguucaucug GUUCGACUGGGACUUCAAGUAG	65.17±4.58	24.1	21.78±1.73	78.0	50.00/50.00 50.00/50.00	50.0/50.0
123 *124R*	gcugacccugaaguucaucugc UUCGACUGGGACUUCAAGUAGA	38.04±4.58	55.7	12.63±1.53	87.2	45.45/54.55 54.55/45.45	50.0/50.0
124 *126R*	cugacccugaaguucaucugca UCGACUGGGACUUCAAGUAGAC	38.03±1.13	55.7	12.65±.056	87.2	50.00/50.00 50.00/50.00	50.0/50.0
131 *132R*	acccugaaguucaucugcacca ACUGGGACUUCAAGUAGACGUG	30.23±3.03	64.8	9.90±1.05	90.0	50.00/50.00 50.00/50.00	50.0/50.0
101 *102R*	cccugaaguucaucugcaccac CUGGGACUUCAAGUAGACGUGG	43.73±0.69	49.1	16.41±0.22	83.4	45.45/54.55 45.45/54.55	45.0/55.0
115 *116R*	ucaucugcaccaccggcaagcu CAAGUAGACGUGGUGGCCGUUC	83.99±0.76	2.2	84.52±5.41	14.7	40.91/59.09 40.91/59.09	40.0/60.0
99 *104R*	caucugcaccaccggcaagcug AAGUAGACGUGGUGGCCGUUCG	80.07±2.04	6.8	63.24±5.29	36.2	36.36/63.64 40.91/59.09	35.0/65.0
109 *110R*	aucugcaccaccggcaagcugc AGUAGACGUGGUGGCCGUUCGA	83.28±1.31	3	86.49±6.09	12.7	36.36/63.64 40.91/59.09	40.0/60.0
129 130R	cugcaccaccggcaagcugccc UAGACGUGGUGGCCGUUCGACG	87.76±0.74	0	73.99±2.76	25.3	27.27/72.73 36.36/63.64	30.0/70.0

observation is consistent with data in lower eukaryotes that suggests that because of the helicase activity of at least one member of the RISC complex any complex secondary structure can be unwound to allow for complete interaction of the RISC with the target mRNA.

Of the 16 small dsRNAs assayed 12 different combinations of two nts covering the putative cleavage site (the middle of the small dsRNA) were identified. However, all but one of the six cleavage dinucleotides present in the small dsRNAs that showed highly effective RNAi were also present in small dsRNAs that mediated an intermediate level of RNAi or that showed no ability to mediate RNAi at all, suggesting that there is no specific internal dinucleotide motif targeted by the small dsRNA. It was not possible to assess if larger sequence motifs are present using the number of dsRNAs tested.

The clearest parameter that correlated with the ability of a dsRNA to mediate highly effective RNAi was the overall percentage of the component pyrimidine and purine bases. Those dsRNAs that included a sense strand with 40.91% or less adenosine (A) and uridine (U) nucleotides failed to mediate RNAi whereas all dsRNAs with a percentage of 45.45% A or U's or higher showed some level of gene expression knock-down. Interestingly the ratio of AU:GC on the antisense strand did not correlate with the ability of a given siRNA to mediate RNAi. These ratios were calculated including the composition of the 3' two nucleotide overhang. Analysis of the core 20 nt dsRNA showed a similar correlation with one exception (SEQ ID NOs: 108 and 109) where a skewed ratio of AU:GC of 40:60 resulted in no inhibition.

Unlike the analysis of the complete dsRNA sequence composition, which should no correlation between the ratio of pyrimidines and purines and the degree of inhibition, a possible association between the AU:GC ratio of the dsRNA 20 nt stretch and the level of RNAi could be seen. Overall, those dsRNAs with a 50:50 AU:GC tended to show a intermediate level of RNAi (one exception, SEQ ID NOs: 116 and 117), and those dsRNAs with a 5% skew away from this ratio (in either direction) tended to mediate a higher degree of inhibition (one exception, SEQ ID NOs:128 and 129).

The effect of dsRNAs against different regions of the marker gene ß–galactosidase (*LacZ*) was also analyzed, in mouse 3T3 cells (Fig. 7). LacZ encodes a much larger transcript than egfp (over 3 kb versus 720 bp) and thus identifying an effective small dsRNA against this transcript may be more difficult. Of the 10 different small dsRNAs tested (Table 3; LacZ Z1-Z10) four mediated a significant degree of silencing of the target gene (Fig. 7A). Interestingly this most effective small dsRNAs clustered in a region of the *LacZ* gene spanning from nts 731 to 1341 and had pyrimidine: purine ratios of 45.00:55.00 through 54.55: 45.45. One small dsRNA, (*LacZ* Z7; SEQ ID NOs: 160 and 161) (50:50 AU:GC) consistently mediated a high degree of inhibition of *LacZ* (Fig 7B).

Effects of the size and modifications of structural and sequence characteristics of the small dsRNA overhang on inhibition on gene expression.

Small interfering RNAs generated *in vivo* in lower eukaryotes by the action of the Dicer RNase III enzyme on large dsRNA molecules have been shown to have short (2 or 3 nt) 3' overhangs. To assess the relative importance of this structural characteristic in synthetic siRNAs, the

ability of small dsRNAs corresponding to region B of *egfp* was initially examined with different 3' and 5' overhangs to mediate inhibition of eGFP expression. Embryonic mouse fibroblasts were co-transfected with a plasmid expressing an enhanced version of GFP (eGFP) and dsRNAs corresponding to *LacZ* (control) and *egfp* (Fig. 8). Cells were harvested 48 hours after initiation of transfection and were assayed for GFP expression by FACS analysis. The results, shown in Fig. 8, demonstrate that small dsRNAs with a variety of different termini can induce RNA. Termini illustrated include a small dsRNA with a two nucleotide 5' overhang, a small double stranded RNA with a two nt 3' overhang on the sense strand and a three nt 3' overhang on the antisense strand, a small double stranded RNA with a three nt 3' overhang on the sense strand and a two nt 3' overhang on the antisense strand, and a small dsRNA with a 0 nt overhang (*i.e.*, blunt end). In this example the small dsRNA with blunt ends shows the highest degree of interference.

To see if small double stranded RNAs can mediate inhibition of pre-existing gene expression human 293 cells stably expressing a modified version of EGFP (d2eGFP) with a protein half-life of approximately 2 hours were transfected twice with small dsRNAs corresponding *LacZ* (control) or *egfp*. The two transfections were conducted 48 hours apart and cells were assayed for GFP expression by FACs analysis 48 hours after the second transfection. This example (Fig 9) shows that a variety of small dsRNAs can interfere with or inhibit *egfp* expression, including dsRNAs corresponding to different portions of the *egfp* sequence, (the 20, 22 and 23 nt dsRNAs correspond to different parts of the *egfp* sequence), and dsRNAs with 3 nt, 2 nt or 0 nt 3' overhangs.

We also assessed the same dsRNAs in HeLa/d2eGFP cells. All of the small dsRNAs with a 3' overhang induced a significant inhibition of eGFP expression, as too did a dsRNA with flush ends. In contrast, though a small dsRNA with a two nt 5' overhang could mediate inhibition the degree of silencing was significantly attenuated (Fig. 10A).

The effects of the orientation and the size of the overhang were further analyzed by studying 137 small dsRNAs generated from combinations of the oligoribonucleotides shown in Table 3 *egfp* region B RNAs #1-32 , corresponding to SEQ ID NOs: 3, 4, 33-36, 97-100, 107-146) . The small dsRNAs tested had 5 or 3' overhangs ranging from 1 to 8 nts, or no overhang at all (Fig. 10B). Of the 137 small dsRNAs studies, 19 (63%) showed highly effective levels of RNAi, of which 12 had 0, 1 or 2 nucleotide overhangs. These data suggest that for efficient interaction of the siRNA with the RISC complex and/or interaction and degradation with the target transcript a dsRNA molecule with either no overhang or a 3' overhang is preferable.

We also assessed whether sequence modifications of the 3' overhang modulates the ability of the region B 22 nt *egfp* dsRNA to induce inhibition of gene expression (Fig. 10C and 10D). Common sequence modifications that have been reported for siRNAs are the replacement of the corresponding endogenous sequence with a deoxythymidine (dTdT) or an uridine (UU) doublet. These changes have been recommended as they can make easier the synthesis of the component oligoribonucleotides. Though the modification of the 3' overhang sequence made no difference to the level of RNAi when eGFP fluorescence intensity was measured, modification of the antisense

strand did reduce the percentage of cells showing complete knockdown of eGFP expression, suggesting a difference in the tolerance of the two strands for such alterations.

Double stranded RNAs incorporating chemical modifications of the component ribonucleosides can mediate interference of gene expression in *C. elegans*.

Several modifications in the phosphate-sugar RNA backbone were assessed for their effects on the ability of dsRNA to trigger interference (Fig. 11A). Modifications of phosphate residues to thiphosphate could be effectively carried out by incorporating thiophosphate nucleotide analogs with T7 and T3 RNA polymerase. Although the α-thiophosphate modifications caused some chemical instability in the RNA, interference activity was demonstrated following incorporation of any single modified residue. Modifications of A, C, or G residues were compatible with full interference activity, while modified U caused some decrease in interference activity (Fig. 11B). Interestingly, Zamore *et al.* (*Cell* 101 25-33, 2000) have noted a preference for U residues in RNA-associated cleavage *in vitro*. RNAs with two modified bases also had substantial decreases in effectiveness as RNAi triggers; modification of more than two residues greatly destabilized the RNAs *in vitro* and we were not able to assay interference activities.

A second position at which modifications were tested was the 2' position of the nucleotide sugar (Fig 12A). Modification of cytidine to deoxycytidine (or uracil to thymidine) on either the sense or the antisense strand of the trigger was sufficient to produce a substantial decrease in interference activity (Fig. 12B). In the case of cytidine to deoxycytidine substitution, this effect must be a consequence of a change at the 2' position, while the effects of uracil to thymidine substitution could reflect effects of the additional methyl group on the thymidine base. Modification of uracil with 2'-fluorouracil was compatible with RNAi activity, while modification with 2'-aminouracil or 2'-aminocytidine produced a decrease in activity comparable to that seen with the deoxynucleotide modification.

A second means to assess requirements at the 2' position involves the question of whether RNA:DNA hybrids can trigger RNAi. Such hybrids were prepared synthetically and enzymatically and found to lack activity (Fig. 12B and data not shown).

Interestingly, a preferential effect on interference activity was observed for several modifications (uracil to 2'-aminouracil, cytidine to 2'-aminocytidine, uracil to thymidine, and cytidine to 2'-deoxycytidine) depending on whether the sense or antisense strand was modified. In each case, trigger activity was more sensitive to modification of the antisense strand than of the sense strand.

A small number of RNA base modifications were studied for their effects on the efficacy of RNA interference reaction. The modifications were chosen based on the commercial availability of nucleoside triphosphates that could be incorporated using T3 or T7 RNA polymerases. Five such base analogs (Figure 11A) were tested; the uracil analogs 4-thiouracil, 5-bromouracil, 5-iodouracil, and 5-(3-aminoallyl)-uracil could be readily incorporated in place of uracil, while inosine was incorporated in place of guanosine. As with the backbone modifications, one area of study was to learn whether there were distinct base requirements for the two strands of the RNAi trigger.

As shown in Figure 12, 4-thio-uracil and 5-bromo-uracil (which were compatible with interference) and inosine (which was compatible but produced a substantial decrease in interference activity) showed no detectable difference in effect between the two strands. By contrast, 5-iodouracil and 5-(3-aminoallyl)-uracil showed remarkable strand-specific effects, with substantially greater effects on the antisense strand.

Whether greater sensitivity to modification for the antisense strand was a general phenomenon or a unique property of the *unc-22A* segment used as a trigger was also tested. Although the sense and antisense strands of the unc-22 had comparable base composition, it was conceivable that some specific feature of the sequence might account for the difference in susceptibility between the two strands. Three other RNAi trigger segments were therefore tested for sensitivity to modification by 5-(3-aminoallyl)-uracil (Figure 12C). For each trigger, (a) full or near full activity with a modified sense strand and (b) greatly reduced activity of the modified antisense strand was observed. A rough quantification of the effect was obtained by carrying out titrations of the active sense-modified RNAs. Minimal differences of 10- to 25-fold were observed between the antisense and sense strands in susceptibility of trigger activity to 5-(3-aminoallyl)-uracil modification (Figure 12C).

There are at least two situations in which a given base modification might block activity of the trigger. The first would involve a specific need for the unmodified base in the function of the trigger RNA, while the second would involve an ability of the modified base to block normal function of the dsRNA. The ability of a 26-mer dsRNA lacking uracil in the antisense strand (Table 4) to function as an interference trigger argues against any specific requirement for chemical groups on uracil. Instead, it seems likely that the large substituents on the 5-position of uracil act by sterically blocking recognition or catalysis at a step in RNAi.

Chemical modification of the component single stranded oligoribonucleotides suggests an asymmetric role for the sense and antisense strands in RNAi in mammalian cells

Chemical modifications including replacement of all of the 2'-hydroxyl by 2'O-methyl (Me) of the component single stranded RNAs used to form dsRNAs have been used to determine the relative importance of the sense and antisense strands in triggering RNA in *C. elegans* and *Drosophila* (Parrish *et al, Mol. Cell* 6 1077, 2000; Elbashir *et al., EMBO J* 20:6877, 2001). In HeLa/d2eGFP cells 2' O-Me residue modification of the antisense stranded abolished mediation of RNAi by a 22 nt *egfp* dsRNA (SEQ ID NO: 138; see table 3 for sequences), however, modification of the sense strand was more tolerated with the degree of inhibition reduced, but still evident suggesting an asymmetric role for the sense and antisense strands (Fig. 13A) in RNAi in mammalian cells.

Small interfering RNAs generated in lower eukaryotes by the Dicer RNase III digestion of larger dsRNAs have been shown to have a 5' Phosphate group. To assess the importance of this in the mediation of RNAi in mammalian cells by small dsRNAs we transfected HeLa/d2eGFP cells with a 22 nt *egfp* dsRNA (SEQ ID NOs: 139 and 140; see table 3 for sequences).where the $5'PO_4$ group was either omitted from both strands or either the sense or antisense strand (Fig 13B). The presence or absence of the 5' phosphate group had no effect on the degree of RNAi seen, suggesting that either

this modifying group is completely unnecessary for the function of the small dsRNAs in mammalian cells or because a phosphate group could be added *in vivo* following transfection into a cell.

To further investigate the effect of modifying the 5' or 3' ends of the dsRNAs we tested the effect of adding a Rhodamine (Rh) fluorophore (TMARA) group to either the 5' (Fig 13C) or the 3' (Fig 13D) end of one or either or both of the ssRNAs (see table 3 for sequences) used to form the dsRNA. These modifications had minimal effect on the ability of a dsRNA to mediate RNAi in the HeLa/d2eGFP cells. However, as seen with the 2' O-Me substitution a slight asymmetric effect was seen with respect to degree of inhibition observed. A dsRNA formed from a 5'Rh or 3'Rh tagged sense strand and unmodified antisense inhibited as effectively as an unmodified duplex, however, a dsRNA formed from an unmodified sense strand and a 5'Rh or 3'Rh tagged antisense strand showed a reduced effectiveness, mediating approximately 50% less inhibition of gene expression compared to an unmodified duplex.

The uptake of small dsRNAs correlates with inhibition of gene expression and though this effect is transient it can be re-established,

Small dsRNAs corresponding to *egfp* and tagged with a Rhodamine (TMARA) group on the 3' end of the sense strand only, were used to follow the uptake of RNA duplexes and to correlate this up-take with RNAi of the *egfp* gene in HeLa/d2eGFP cells. Using fluorescent images of cells taken with an AxioPlan2 microscope at 25 fold magnification small dsRNAs up-take could be seen with 6 hours after initiation of transfection. The uptake of small dsRNA was maximal at 48 hours. Importantly, a decrease in eGFP expression was also first noted 6 hours after initiation of the transfection and the peak level of inhibition of eGFP expression was also seen at 48 hours. The inhibition of eGFP expression was studied over a longer period (12 days) and RNAi against this gene was shown to peak at 48 hours post transfection and had returned to pre-treatment levels by day 12 (Fig 14). However, re-treatment of Hela/d2eGFP cells with the same dsRNAs completely re-established RNAi against the target gene and to a similar level to that seen following the first transfection. This data shows that, though RNAi is transient in HeLa cells, and thus probably in most mammalian cells, repeated administration of the dsRNA permits longer-term silencing of the particular target gene in question.

Variations in the sequence of the small dsRNA

The sequences of the small dsRNAs in this disclosure are substantially identical to the sequence of the target RNA for which interference or inhibition of expression is desired. For purposes of this disclosure, the sequence of the RNA is "substantially identical" to a specific portion region of the target RNA if it differs by no more than about 30 percent, and in some embodiments no more than about 10 percent, from the specific portion of the target RNA. However, in some cases identity with the target RNA may not be as exact, as in the case of targeting a transcript with polymorphic or mutant sequences or alternative spliced forms. In these situations there may be a need to either target all of the alternative sequence forms of the particular transcript, or there may be a need to specifically distinguish between these transcripts.

To illustrate that small dsRNAs with non-identical sequences can target a transcript in a manner that may allow allelic discrimination or the general targeting of a transcript that is encoded by the same gene but which has multiple forms, small dsRNAs against *egfp* were assessed where one, or either, or both strands of the dsRNA contained a three nucleotide change from the primary target sequence (see table 3 for sequences and Fig. 15). As with other modifications of the component strands of the dsRNA we saw an asymmetric effect of the three nucleotide change. When the there nucleotide change was present on the sense strand there was no effect on the ability of the dsRNA to mediate RNAi, however, when on the antisense strand the percentage of cells inhibited was significantly less. This suggests that small (between 1-10 nt) variations in sequence could be used to distinguish between transcripts or to ensure that all of the sequence variants of a particular transcript are targeted. The three nucleotide change used here was incorporated in approximately the middle of the dsRNA sequence, but variations in the sequence could be incorporated into any position within the dsRNA.

Example 3: Gene Silencing Microarrays

This example describes one type of gene silencing microarray, in which dsRNAs are transfected into mammalian cells from a microarray format, and the effects of gene silencing by RNAi are studied.

Methods

Nucleic acids. Single stranded RNA oligonucleotides were chemically synthesized and HPLC purified by Xeragon Inc. (Huntsville, AL; Gaithersburg, MD). Duplex RNA molecules were generated as described in Example 2. The integrity of the dsRNAs was confirmed by gel electrophoresis. To assess cellular localization of dsRNAs an *egfp* dsRNA was synthesized with a rhodamine (TAMRA) (*rh-egfp* dsRNA) added to the 3' end of the sense strand of the RNA duplex; this addition has no effect on the efficacy with which the dsRNA mediates gene silencing (Example 2; Fig. 13D). The sequences of the RNA oligonucleotides used to form the dsRNAs used in this Example are:

cat sense 5'PO_4r(gagugaauaccacgacgauuuc)3' (SEQ ID NO: 15);

cat antisense 5'PO_4r(aaucgucgugguauucacucca)3' (SEQ ID NO: 16);

egfp sense 5'PO_4r(gcaagcugacccugaaguucau)3' (SEQ ID NO: 3);

egfp antisense 5'PO_4r(gaacuucagggucagcuugccg)3' (SEQ ID NO: 4);

egfpB sense 5'PO_4r(cccugaaguucatcugcaccac)3' (SEQ ID NO: 97);

egfpB antisense 5'PO_4r(ggugcagaugaacuucagggtc)3' (SEQ ID NO: 98);

egfpC sense 5' PO_4r(caucugcaccaccggcaagcug)3' (SEQ ID NO: 99); and

egfpC antisense 5'PO_4r(gcuugccgguggugcagaugaa)3' (SEQ ID NO: 100).

Plasmid pEGFPd2-N3 (Clontech, Palo Alto, CA) expresses a mammalian enhanced version of green fluorescent protein (eGFP) with a defined protein half-life of two hours.

Cell culture. HeLa cells, a human epithelial cell line derived from a cervical adenocarcinoma (American tissue culture collection (ATTC), Manassas, VA), were grown in Dulbeccos Modified Eagles Medium (Life Technologies, Gaithersburg, MD) supplemented with 10% fetal bovine serum (FBS, Gemini BioProducts Inc., Calabasas, CA) (D10) at 37° C/5% CO_2. HeLa/d2eGFP cells stably expressing the destabilized version of EGFP were generated by cationic lipid transfection of the pd2EGFP-N3 (Clontech) plasmid and selection with 1 mg/ml G418. To obtain a population of GFP positive cells with a narrow range of fluorescence intensity, cells were subjected to flow cytometric sorting.

RNAi microarray. For transfection on a microarray platform, a lipoplex of dsRNA (0.25 μg to 2.5 μg) and lipid (Lipofectin, Life Technologies; 1:2 wt:wt RNA:lipid) was formed, at room temperature, over a period of 15 minutes in OpitMEM (Life Technologies) supplemented with 0.2M sucrose in a final volume of 100 μl. After formation of the lipoplex, 0.5-1 μl of the basement membrane matrix Matrigel™ (BD Biosciences Discovery Labware, Bedford, MA) was added, and between 0.1 and 1 μl (0.25 ng – 2.5 ng dsRNA) of the mixture was arrayed on to poly-lysine coated plates (Mousses *et al.*, in *Functional Genomics*. (ed. F.J.H. Livesey, S.P) 113-137, Oxford University Press, Oxford; 2000). dsRNA spots between 100-500 μM in diameter and with a center-center distance of 300-1000 μM apart have been successfully arrayed and analyzed.

Slides were pre-treated for 1 hour in OptiMEM at 37° C/5% CO_2 to remove surface dsRNA. Slides were transferred to fresh sterile dishes (10cm^2) and cells in OptiMEM medium (10 ml) were plated onto of the slides at a concentration of ~5x10^5 cells per ml. Approximately 18 hours after cell plating, an equal volume of growth medium supplemented with 20% FBS was added. Cells were assayed 24 to 72 hours after the plating of cells.

Imaging, quantification, and statistical analysis. An AxioPlan2 microscope fitted with a Marchauser 8 position stage and an Axiocam CCD digital camera (Carl Zeiss GmbH, Oberkochen, Germany) at 25 to 1000-fold magnification was applied for automated image acquisition (100-200 fold magnification) using a custom macro written in the scripting language of the KS400 (Zeiss) software package. A multi-step process was used to generate binary images for measuring fluorescence intensities in each channel for segmented cells.

The steps in the image analysis for the current embodiment were: (1) the DAPI (blue channel) images and the green image were regionally thresholded to generate binary images. (2) These two binary images were combined to produce a merged image of the union of the two thresholds. (3) The green image was analyzed to determine the graduations in the green signal intensity gradients and 'watershed' lines were placed at the inflection points in the valleys of this green signal using the DAPI derived binary image as a marker. (4) The merged binary image in step 2 was now combined with the green image subjected to water-shedding to locate and define the precise area of every cell in the image. (5) Partial cells and noise were subtracted to produce a fully segmented image suitable for extraction of multi-parametric data.

Statistical analysis of the multiparametric data including generation of global statistics of cell populations in each image was performed automatically with the help of MATLAB 6 (The

MathWorks, Inc., Natick MA). Local background was calculated and subtracted from the global statistics. All statistical tests in this study were performed using a two-tailed Student t-Test. An overview of a representative embodiment of the RNAi microarray system, including a web-based information management system used to coordinate various aspects of the overall system, is illustrated in Figure 16.

Results and Discussion

Described herein is a novel microarray-based method for parallel analysis of multiple RNAi events in cultured human cells. In the currently described embodiment, dsRNAs were arrayed on glass slides, overlaid with a monolayer of cells, and gene silencing assessed by quantitative microscopic image analysis at a single cell level. RNAi in a microarray format provides potent, sequence specific and spatially confined gene silencing, enabling cell-based analyses of gene function on a genome scale.

In order to replicate the generation of mammalian "somatic knockout" phenotypes using RNAi in a microarray format, dsRNAs corresponding to the enhanced Green Fluorescent Protein (*egfp*) gene were arrayed on microscope slides, and plated HeLa cells permanently expressing eGFP (HeLa/d2eGFP cells) on top of the slides. The uptake of *rh-egfp* dsRNA led to a spatially confined silencing of the GFP expression, as visualized by 'black holes' in the monolayer of the green HeLa/d2eGFP cells, the affected cells within the *rh-egfp* dsRNA spot were visualized by DAPI staining and showed no indication of cytotoxicity. Furthermore, red fluorescence indicating intracellular accumulation of *rh-egfp* dsRNA was observed in these affected cells.

The dsRNAs were printed in a cationic lipid/Matrigel mixture, which helped to confine the spread of the dsRNA, leading to a clearly defined edge to the spot, virtually complete inhibition of green fluorescence within the spot and little or no inhibition surrounding the spot. Together, these observations show that arrayed dsRNAs on a glass slide enter cells via transfection, and mediate potent inhibition of the target protein expression in a spatially confined manner.

The assessment of the functional effects of RNAi in a microarray format benefits from a capability for quantitative image analysis, with single cell resolution, for thousands of "spots" on an array. Image analysis at the spot level using microarray laser scanners has insufficient resolution (maximum resolution of 5 μm) for cellular, and sub-cellular cytometry. A rapid, automated, high-resolution, microscope based image acquisition system was therefore developed for RNAi microarrays, as was an algorithm for quantitative image analysis.

In the current example, sets of three-color images were acquired: the blue channel was used for DAPI counter-stain (cell enumeration and defining nuclear compartment in cells), the green channel for a phenotypic marker (such as for eGFP expression in this study), and the red channel for a second marker (such as uptake of rhodamine tagged dsRNAs in this study). Cells in the image were segmented, and multiple parameters were extracted from each individual cell (including the total, minimum, maximum, mean, and median fluorescence intensity).

Summary statistics for these parameters were calculated for the population of cells within each feature. These parameters can be extracted for nuclear, cytoplasmic cell areas from the same microscopic field at several different wavelengths. In this example, the percentage of GFP positive HeLa/d2eGFP cells declined from 79% (in the negative control spot) to 19% (*egfp* dsRNA spot).

To demonstrate the sensitivity and specificity of RNAi microarray and automated image analysis we conducted the following experiments. First, a significant time-dependent reduction in eGFP fluorescence was seen ($p < 0.001$) in spots containing the *rh-egfp* dsRNA (Fig. 16A).

Second, a modest effect of dose, consistent with previous observations by other methods (Example 2), was observed at both time points within *rh-egfp* dsRNA spots, but not within the negative control dsRNA spots (Fig. 16B and C).

Third, other parameters, such as mean cell areas (Fig. 16D), were not affected by either of the two dsRNA.

Fourthly, three dsRNAs corresponding to different sequences within the *egfp* gene showed differential effectiveness in reducing eGFP expression (Fig. 16E).

Finally, the RNAi microarray results for these three *egfp* dsRNAs produced similar results to those obtained using parallel flow cytometry measurements (Fig. 16F). Given this variability in efficacy of different dsRNAs targeting the same gene the data presented here demonstrates that RNAi microarrays can also be used as a platform for optimization of dsRNA design.

This example clearly demonstrates the sequence, time, and dose dependent inhibition of green fluorescence mediated by the *rh-egfp* dsRNA and validates RNAi on a microarray platform as a means of studying gene silencing in cultured mammalian cells. In addition the image analysis system and web-based image analysis and database can be applied to not only the RNAi microarray system but to cells grown on chamber slides (See Examples 5 and 6) and in plates, including, though not limited to, 96 well microtiter plates.

Example 4: Rescue of Polyglutamine Mediated Cytotoxicity by Double-Stranded RNA-Mediated RNA Interference

In this example, *Drosophila* and human tissue culture models of the dominant genetic disorder spinobulbar muscular atrophy (SBMA) are used to determine if RNAi can be used to specifically down-regulate a human disease related transcript. A variety of different dsRNAs were assessed for the ability to inhibit expression of transcripts that included a truncated human androgen receptor (*ar*) gene containing different CAG repeat lengths (16-112 repeats). In *Drosophila* cells, dsRNAs corresponding to non-repetitive sequences mediated a high degree of sequence-specific inhibition, while RNA duplexes containing *cag* repeat tracts only induced gene-specific inhibition when flanking *ar* sequences were included; dsRNAs containing various lengths of *cag* repeats plus *ar* sequences were unable to induce allele-specific interference. In mammalian cells, sequence-specific small dsRNAs of 22 nts were tested; these rescued the toxicity and caspase-3 activation induced by plasmids expressing a transcript encoding an expanded polyglutamine tract. This example

demonstrates the feasibility of targeting a transcript associated with an important group of genetic diseases by RNAi.

Methods

Plasmids and generation of dsRNA. The construction of pAct.GFP.and pAct.CAT have been previously described (Caplen *et al.*, *Gene* 252:95-105, 2000). The plasmids pAct.ARCAG26GFP, pAct.ARCAG43GFP and pAct.ARCAG106GFP were constructed by insertion of a previously constructed cassette containing 178 bp (nt 524 to nt 702) and 207 bp (1345 nt to 1552 nt) of the human androgen receptor 5' and 3' of 112 CAG repeats in frame with *egfp* (Clontech, Palo Alto, CA) down stream of the *Drosophila* actin 5c promoter in pRact.Adh (L. Cherabas, Indiana University, IN). Clones were screened for contraction or expansions of the CAG repeat during cloning and three clones with 26, 43 and 106 CAG trinucleotide repeats were identified (Fig. 18A and 18B). The $pAR_{(N)}$ constructs, pCMV-ARCAG16 and pCMV-ARCAG112 have been described previously (Merry *et al.*, *Hum. Mol. Genet.* 7:693-701, 1998). The pCMV-ARCAG19GFP and pCMV-ARCAG112.GFP plasmids were generated by sub-cloning *Xba*1/*Eco*R1 fragments from pAR(N) into pEGFP-N1 (Clontech) generating an in-frame fusion cDNA consisting of a truncated form of the human androgen receptor and enhanced green fluorescent protein (EGFP).

Single stranded, gene specific sense and antisense RNA oligomers corresponding to the ß-galactosidase (*LacZ*) gene, the chloramphenicol acetyl transferase (*cat*) gene, the green fluorescent protein (*gfp*) gene and, the human androgen receptor (*ar*) gene (sequences shown in Table 5) were synthesized using 2'-O-(triisopropyl)silyloxymethyl chemistry by Xeragon AG (Zurich, Switzerland). Sense and antisense oligoribonucelotides consisting of 27 CAG trinucleotides and 27 CAG trinucleotide repeats plus 21 nts from the 5' region of the human androgen receptor immediately adjacent to the CAG tract (see Table 5 for sequences) were also chemically synthesized. To form dsRNA molecules 100 μg of each complementary ssRNA oligomer in 10 mM Tris-Cl (pH 7.0), 20 mM NaCl (total volume 300 μl) was heat denatured at 95° C in a water bath and then allowed to anneal slowly (in the water bath) for approximately 18 hours. The dsRNAs were ethanol precipitated and re-suspend in RNase free water at a concentration of approximately 0.5 mg/ml. Double stranded RNA molecules corresponding to *arcag26ar* and *arcag43ar* and *arcag106ar* were generated by sub-cloning each cassette as an *Eco* RI/*Bss* HII fragment into pCMV-BK (Statagene Inc.). Sense and antisense RNAs were synthesized for each cassette using standard T3 and T7 RNA polymerase *in vitro* transcription reactions (Megascript, Ambion) after preparation of a *Pvu* II/*Nhe* I fragment which included non-homologous plasmid sequences. Double stranded RNAs were generated as above; after precipitation flanking ssRNAs derived from non-homologous plasmid sequences were removed using Ribonulcease T1 in 0.3 M NaCl. The integrity and change in mobility of all of the dsRNAs were confirmed by gel electrophoresis. The position of each of the dsRNAs relative to its target is shown in Fig. 18C.

Table 5. Sequences of dsRNAs used in Example 4

SEQ ID NO:	ssRNA	Sequence
45	*lacZ* s 81 nts	5'r(agcuccugcacuggauggcggcgcuggaugguaagccgcuggcaagcggugaagugcc ucuggaugucgcuccacaaggua)3'
46	*lacZ* as 81 nts	5'r(uaccuuguggagcgacauccagaggcacuucaccgcuugccagcggcuuaccauccagc gccaccauccagugcaggagcu)3'
41	*cat* s 78 nts	5'r(caucgcucuggagugaauaccacgacgauuuccggcaguuucuacacauauauucgcaa gauguggcguguuacggug)3'
42	*cat* as 78 nts	5'r(caccguaacacgccacaucuugcgaauauauguguagaaacugccggaaaucgucgugg uauucacuccagagcgaug)3'
39	*gfp* s 78 nts	5'r(cuacguccaggagcgcaccaucuucuucaaggacgacggcaacuacaagacccgcgccga ggugaaguucgagggcga)3'
40	*gfp* as 78 nts	5'r(ucgcccucgaacuucaccucggcgcgggucuuguaguugccgucguccuugaagaagau ggugcgcuccuggacguag)3'
168	*ar* s 78 nts	5'r(agggugaggaugguucucccaagcccaucguagaggccccacaggcuaccuggucug gaugaggaacagcaaccuu)3'
169	*ar* as 78 nts	5'r(aagguugcuguuccucauccaggaccagguagccuguggggccucuacgaugggcuug gggagaaccauccucacccu)3'
170	*cag27* s 81 nts	5'r(cagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag cagcagcagcagcagcagcag)3'
171	*cag 27* as 81 nts	5'r(cugcugcugcugcugcugcugcugcugcugcugcugcugcugcugcugcugcugcugcugc ugcugcugcugcugcugcug)3'
172	*arcag27* s 104 nts	5'r(ggcgccaguuugcugcugcugcagcagcagcagcagcagcagcagcagcagcagcagca gcagcagcagcagcagcagcagcagcagcagcagcagcagca)3'
173	*arcag27* as 104 nts	5'r(ugcugcugcugcugcugcugcugcugcugcugcugcugcugcugcugcugcugcugcu gcugcugcugcugcugcugcugcagcagcagcaaacuggcgcc)3'
15	*cat* s 22 nt	5'PO_4r(gagugaauaccacgacgauuuc)3'OH
16	*cat* as 22 nt	5'PO_4r(aaucgucgugguauucacucca)3'OH
3	*gfp* s 22nt	5'PO_4r(gcaagcugacccugaaguucau)3'OH
4	*gfp as* 22nt	5'PO_4r(gaacuucagggucagcuugccg)3'OH
174	*5' ar* s 22 nts	5'PO_4r(gcucaaggauggaagugcaguu)3'OH
175	*5' ar* as 22 nts	5'PO_4r(cugcacuuccauccuugagcuu)3'OH
176	*3' ar* s 22 nts	5'PO_4r(cuggaugaggaacagcaaccuu)3'OH
177	*3' ar* as 22 nts	5'PO_4r(gguugcuguuccucauccagga)3'OH

Single stranded oligoribonucleotides used to form double stranded RNAs. (s = sense, as = antisense).

Cell culture and nucleic acid transfections. S2 cells were grown in DES™ Medium (Invitrogen) supplemented with 10% fetal bovine serum (FBS). Cells were passaged every 2 to 3 days to maintain exponential growth. S2 Cells were transfected using either the cationic lipid CellFectin (Life Technologies) using an adaptation of the manufacturer's protocol. Briefly, cells were seeded and allowed to settle overnight, nucleic acid (plasmid and/or dsRNA) was complexed with lipid at a weight to weight ratio of 1:6 in DES medium without supplementation. The complex was incubated at room temperature for 15 minutes and then added to cells from which normal growth medium had been removed. After overnight incubation an equal volume of DES medium plus 20% FBS was added to the cell / lipoplex mixture. HEK-293T cells were maintained in DMEM medium (Gibco BRL) supplemented with 10% FBS. Cells were plated at a density of 60% in 6-well culture dishes the day prior to transfection. Transfection with 1 μg DNA was carried out using Lipofectamine Plus following the manufacturers protocol (Gibco BRL) plus 0.5 μg dsRNA.

Transgene expression. GFP expression was assessed by FACs analysis or deconvolution microscopy. For FACs analysis (FacsCaliber, Becton Dickinson) pAct.CAT transfected cells were used to gate for forward scatter and side scatter. The percentage of GFP positive cells was

determined by gating against pAct.CAT transfected cells with M1 falling between 10 or 20 and 9910 (FL-1), the geometric mean was used as a measure of the relative intensity of fluorescence. 10,000 non-gated events were acquired for each sample. For deconvolution microscopy S2 and 293 cells were either spun onto slides or grown on coverslips or 2-well chamber-slides (Nunc). At the indicated times the cells were fixed in 4% paraformaldehyde for 15 minutes at room temperature then washed three times in PBS. All cells were stained with 1 μg/ml Hoechst 33342 (Sigma) in PBS for 10 minutes. Deconvolved images were produced on an Olympus microscope using a 60X water-immersion objective and Deltavision software (Applied Precision) on a Silicon Graphics workstation. Multiple optical sections 0.2 μm thick were analyzed.

To directly measure the toxicity of expanded polyglutamine in the context of truncated androgen receptor we employed a FACS-based survival assay. HEK-293 cells were harvested with trypsin, gently pelleted by centrifugation and resuspended in PBS with 0.5% serum ice at a concentration 10^6/ml. The cells were stained with propidium iodide 1 μg/ml (PI, Sigma), gently vortexed and incubated for 15 minutes at room temperature in the dark. 50,000 non-gated events were acquired for each sample (Beckman Coulter XL instrument and software package used for analysis). Results are expressed as a percentage of PI-negative (viable) cells (FL-2 channel) relative to total GFP positive (transfected) cells (FL-1 channel).

Chloramphenicol acetyl transferase expression was assessed using an ELISA based assayed (Roche Biochemicals). Total protein was determined using the Bradford method and the micro-assay protocol adapted for use in 96 well micro-titer plates (Bio-Rad). Absorption readings (A_{595}) were converted to absolute amounts using a bovine serum albumin (Sigma) (0.625 -10 μg) standard curve after subtraction of background values. Statistical analyses were conducted using an unpaired t-test (SatView) for comparison of groups. The null hypothesis was rejected at $p>0.05$.

RNA analysis. Total RNA was isolated from using GTC extraction and poly(A+) RNA selected using Oligo dT cellulose (Ambion, Inc., Austin, TX). The poly(A+) RNA was subjected to DNAse digestion to remove any trace contamination of plasmid DNA and analyzed by electrophoresis (1.2 % agarose, 1xMOPS, 5.0% formaldehyde), Northern blot transfer and hybridization (Ambion) at 42°C with ^{32}P–labeled random-primed DNA probes. A human androgen receptor gene specific probe (148 bp) was prepared by PCR amplification using pCMV.AR112GFP as template and the following oligonucleotides (*ar* forward: 5' ggaagtgcagttagggctg 3' and *ar* reverse 5' cagcagcagcaaactggcg 3' SEQ ID NOs: 313 and 314). The *gfp* and *Drosophila Gapdh-1* DNA probes have been previously described (Caplen *et al.*, *Gene* 252:95-105, 2000). In all cases, filters were washed at high stringency (0.2 X SSC) and subjected to autoradiography. The intensities of the hybridization signals were obtained using a Fuji BAS1500 phosphoimager (Fuji Photo Film Co. Ltd., Tokyo, Japan) and pixel densities calculated using Image reader 1.4 and ImageGauge 3.0 (Fuji Photo Film Co. Ltd., Tokyo, Japan).

Caspase-3 activity. To quantitate caspase-3 activity, 10^6 cells were washed with cold PBS and harvested by scraping at the indicated times post-transfection. The cells were pelleted at 400 *g* for 5 minutes then re-suspended in lysis buffer (10 mM Tris, pH 7.3, 10 mM NaH_2PO_4, 150 mM

NaCl, 1% Triton X-100). Protein concentration was determined using the DC protein assay reagent (BioRad). 100 mg of cell extract was incubated with 50 mM of fluorometric substrate DEVD-AFC (7-amino-4-trifluoromethylcoumarin, ApoTarget kit, Biosource International) in a total volume of 100 μl of reaction buffer (10 mM Hepes, pH 7.4, 50 mM NaCl, 0.5 mM EDTA, 0.1% CHAPS, 10% glycerol, 10 mM DTT) in the dark for one hour at 37°C. DEVDase activity was determined by measuring the liberation of AFC using a Cytofluor II multiwell plate reader (PerSeptive Biosystems) with excitation and emission wavelengths of 420 and 520 nm, respectively. Caspase-3 specificity was established by blocking activity with the inhibitor DEVD-CHO at 10 mM.

Results

Gene specific dsRNAs can interfere with expression from arcag$_n$gfp fusion transcripts in Drosophila S2 cells

Drosophila embryonic S2 cells have been used extensively as a cell culture system to analyze the RNAi mechanism and its application (Caplen *et al.*, *Gene* 252:95-105, 2000; Clemens *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* 97:6499-6503, 2000; Hammond *et al.*, *Nature* 404:293-296, 2000). Though several transgenic *Drosophila* models of polyQ diseases have been developed, (Marsh *et al.*, *Hum. Mol. Genet.* 9:13-25, 2000; Jackson *et al.*, *Neuron* 21:633-642, 1998; Warrick *et al.*, *Cell* 93:939-949, 1998; Bonini, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 354:1057-1060, 1999), no cell culture model has been described. To develop a *Drosophila* cell culture model of SBMA, *Drosophila* expression plasmids were constructed to express fusion transcripts containing portions of the human androgen receptor (*ar*) gene, with CAG tracts of 26, 43 or 106 repeats and the marker gene green fluorescent protein (*gfp*) (Fig. 18A). Northern analysis of mRNA purified from S2 cells transfected with each expression plasmid showed appropriately sized *gfp* containing transcripts (Fig. 19F).

Transient transfection of S2 cells with plasmids containing 43 and 106 CAG trinucleotide repeats showed GFP positive protein aggregates and aggresomes (Fig. 18B). Low-level GFP associated protein aggregation was observed in S2 cells transfected with pAct.ARCAG43GFP after approximately 2-3 days; localized aggresomes were seen in 4-5 days. S2 cells transfected with pAct.AR.CAG106GFP formed GFP positive protein aggregates and aggresomes within 24 hours. Of those S2 cells transfected, approximately 4% of pAct.ARCAG43GFP and 35% of pAct.ARCAG106GFP transfected cells developed GFP positive protein aggregates 3 days after transfection, this increased to 10% of pAct.ARCAG43GFP transfected cells and 90% of pAct.ARCAG106GFP transfected cells 5 days after transfections. These results are consistent with observations in mammalian cells, where trinucleotide length has been shown to correlate with the degree of protein aggregation. Trinucleotide repeats in the range of 40-50 produce proteins that are susceptible to aggregation but this develops over a longer period of time than when a larger repeat expansion is used, represented in this study by the 106 CAG repeat construct (Butler *et al.*, *Hum. Mol. Genet.* 7:121-127, 1998; Merry *et al.*, *Hum. Mol. Genet.* 7:693-701, 1998; Martindale *et al.*, *Nat. Genet.* 18:150-154, 1998). No protein aggregation was seen when the control pAct.GFP or pAct.ARCAG26GFP plasmids were used.

To assess the effect of gene specific dsRNAs on the expression of the fusion transcripts from each of the pAct.ARCAG$_n$GFP plasmids dsRNAs of approximately 80 nts corresponding to portions of the *gfp* gene and the *ar* gene were generated using the appropriate sense and antisense oligomers (Table 5 and Fig. 18C); dsRNAs corresponding to the ß-galactosdase gene (*LacZ)* and the chloramphenicol acetyl transferase gene (*cat)* were used as controls. *Drosophila* S2 cells were co-transfected with each dsRNA molecule and each of the *Drosophila* expression plasmids. Cells were harvested 72 hours after initiation of the transfection and assayed for GFP expression (Fig. 19). In all cases the *gfp* dsRNA specifically inhibited expression of GFP, including GFP expression from all three of the pAct.ARCAG$_n$GFP fusion plasmids. The *gfp* dsRNA inhibited GFP expression from pAct.ARCAG26GFP by approximately 75%, GFP expression from pAct.ARCAG43GFP by approximately 80% and GFP expression from pAct.ARCAG106GFP by approximately 85%. The 80% inhibition of GFP expression from the pAct.GFP plasmid is consistent with our previous data (Caplen *et al.*, *Gene* 252:95-105, 2000).

The *ar* dsRNA specifically inhibited GFP expression from the fusion transcripts containing human androgen receptor sequences. The *ar* dsRNA induced significant (60-80%) down-regulation of GFP expression from all three of the pAct.AR(CAG)$_n$GFP fusion plasmids (Figs. 19A-19C). The *ar* dsRNA had no effect on GFP expression from the parental plasmid pAct.GFP, which contains no *ar* sequences (Fig. 19D) and only the dsRNA corresponding to *cat* inhibited CAT expression, the *ar* dsRNA had no significant effect on CAT expression (Fig. 19E).

Northern analysis of mRNA purified from *Drosophila* cells treated with these dsRNAs also demonstrate gene specific inhibition with a correlation between the number of *gfp* and *ar* containing transcripts and the level of GFP expression observed (Fig 19F). Using phosphoimage analysis, the levels of the *gfp* positive and *ar* positive signals were first normalized using the *Drosophila* glyceraldehyde 3-phosphate dehydrogenase-1 (*gapdh-1*) hybridization signal; the relative levels of the *ar*/*gfp* positive transcripts were then compared. The *gfp* dsRNA induced a 70-90% decrease in the *ar*/*gfp* positive transcript. The *ar* dsRNA induced a 40% decrease in the level of the *arcag26gfp* transcript; the *arcag43gfp and arcag106gfp* mRNA levels were reduced by 60-80% relative to controls.

The effect of dsRNAs containing cag repeats on transgene expression in Drosophila S2 cells

Several studies of RNAi in invertebrate systems have considered the size effects and sequence content of the input dsRNA used to mediate this process (Caplen *et al.*, *Gene* 252:95-105, 2000; Parrish *et al.*, *Mol. Cell* 6:1077-1087, 2000; Zamore *et al.*, *Cell* 101:25-33, 2000; Tuschl *et al.*, *Genes. Dev.* 13:3191-3197, 1999; Yang *et al.*, *Curr. Biol.* 10:1191-1200, 2000), but a transcript containing different numbers of CAG repeats flanked by complex human sequences represents a novel target. To examine the effect of dsRNAs containing highly repetitive sequences, we used dsRNA consisting of 27 CAG repeats (*cag27*) and 27 CAG repeats plus 21 nts from the 5' end of the human androgen receptor sequence (*arcag27*) (Table 5 and Fig. 18C). Unlike the specific effect seen with the use of complex gene-specific dsRNAs such as the *ar* and *gfp* dsRNAs, the *cag27* dsRNA induced a non-specific inhibition of transgene expression in *Drosophila* cells: Representative data

shown in Figure 4 shows the significant inhibition (60 to 75%) of GFP expression from the pAct.ARCAG$_n$GFP fusion plasmids (Fig. 20A-20C), however, we also consistently saw a decrease in the level of GFP expression from the parental pAct.GFP plasmid of the order of 30 to 40% (Fig. 20D). We tested a number of preparations of the *cag27* dsRNA, but all had the same effect on GFP expression from pAct.GFP. The addition of 21 nts of the *ar* sequence immediately upstream of the CAG repeat did not improve the specificity of this inhibition, as the *arcag27* dsRNA also reduced GFP expression from the pAct.GFP plasmid by approximately 50% (Fig. 20D). GFP expression from the fusion transcripts was inhibited by 70-80% by the *arcag27* dsRNA (Fig. 20A-20C). Interestingly, this non-specific interference with transgene expression, in large part, could be alleviated by the use of larger regions of *ar* sequence on both sides of the CAG repeat tract. This restoration of specificity was irrespective of the size of the intervening CAG tract (Fig. 20E-20H). However, though these dsRNAs showed sequence –specificity, in that they had minimal effect on transgene expression from the parental pAct.GFP plasmid we saw no evidence for allelic-specificity , that is, dsRNAs with different CAG repeat lengths (26, 43 or 106) interfered with expression from all three of the *arcag$_n$gfp* transcripts irrespective of the number of repeats in the target transcript (Fig. 20E-20G). The interference mediated by the *arcag$_n$ar* dsRNAs was of a similar level to that seen using the 78 nt *gfp* dsRNA (Fig 20E-20H).

Rescue of cell toxicity induced by transcripts expressing expanded polyQ repeats by small dsRNAs in human cells

In mammalian cells, dsRNAs of over approximately 80 nts can trigger a variety of cellular responses that result in a non-specific decrease in gene expression and frequently cell death (Clemens and Elia, *J. Interferon Cytokine Res.* 17:503-524, 1997). Recently, however, we and others have shown that small dsRNAs of ~21-23 nts can mediate a sequence-specific inhibition of gene expression in mouse and human cells (Elbashir *et al.*, *Nature* 411:494-498, 2001; Caplen *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* 98:9742-9747, 2001). To determine whether small dsRNAs can specifically interfere with expression of a polyQ protein and thus rescue the increased cell death associated with an expanded CAG tract we co-transfected human HEK-293 cells with mammalian expression plasmids expressing a fusion transcript consisting of a truncated version of the human androgen receptor containing either 112 or 19 CAG repeats and *gfp* (pCMV.ARCAG112GFP and pCMV.ARCAG19GFP respectively) and 22 nt dsRNAs corresponding to either *cat* or *gfp* (Table 5 and Fig.18C). The small dsRNAs used were designed with a 5'PO_4 and 2 nucleotide 3' overhangs, a structure that is characteristic of the RNase III (Dicer) enzymatic cleavage that has been associated with RNAi (Bernstein *et al.*, *Nature* 409:363-366, 2001). We assayed for GFP expression by deconvolution microscopy (Fig. 21A-21F) and assessed toxicity using a FACs based viability assay cell death (Fig. 21G and 21H) 96 hours after initiation of transfection. Consistent with previous observations, (Caplen *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* 98:9742-9747, 2001) a 22 nt *gfp* dsRNA inhibited GFP expression from both fusion transcripts (Fig. 21E and 21F); a 22 nt dsRNAs corresponding to *cat* had no effect on GFP expression. Further, protein aggregates were clearly evident in pCMV.ARCAG112GFP/*cat* 22 nt dsRNA co-transfected cells whereas there was little or

no evidence of protein aggregation in those cells transfected with pCMV.ARCAG112GFP and the 22 nt *gfp* dsRNA. Interference with expression from the ARCAG112GFP transcript mediated by the 22 nt *gfp* dsRNA also significantly reduced the cell death associated with the presence of a polyQ expansion (Fig. 21G) towards control levels (Fig. 21H).

We next tested two 22 nt dsRNAs corresponding to different regions of the human *ar* gene; one region was 5' of the CAG repeat (*5'ar*) and the second target region was 3' of the CAG repeat (*3' ar*). Both dsRNAs significantly reduced the cytotoxic effect mediated by the presence of an AR protein with an expanded polyQ tract (Fig. 21I), with the 22 nt *5' ar* dsRNA inducing the greatest level of rescue by reducing the percentage of dead cells to the range seen in control transfections (Fig 21J). Importantly, both 22 nt *ar* dsRNAs also significantly reduced the levels of Caspase-3 activity induced by the ARCAG112 protein. The control *cat* 22nt dsRNA had no effect on caspase-3 activity indicating the specificity of the effect mediated by the *ar* 22 nt small dsRNAs.

Discussion

Double stranded RNA-mediated silencing of gene expression is increasingly being seen as an important epigenetic mechanism controlling gene expression and as a reverse genetics tool (Kuwabara and Coulson, *Parasitol. Today* 16:347-349, 2000; Vance and Vaucheret, *Science* 292:2277-2280, 2001; Matzke *et al.*, *Science* 293:1080-1083, 2001; Barstead, *Curr. Opin. Chem. Biol.* 5:63-66, 2001). With the recent description of an RNAi-like mechanism in mammalian somatic cell, it is now also possible to exploit dsRNA triggered RNAi as a therapeutic approach akin to current strategies that exploit single stranded antisense oligonucleotides or transcripts to down-regulate gene expression.

This example assessed the feasibility of such an approach in a model system where RNAi has been relatively well characterized by developing a cell culture model of the dominant genetic disorder SBMA (Kennedy's disease) in *Drosophila* S2 cells. The recently described process of small dsRNA (21-23 nt)-triggered RNAi in mammalian cells was employed to assess whether this could be used to rescue the toxic effects of a truncated human androgen receptor protein containing an expanded polyQ tract. In summary, dsRNA-mediated gene-specific, but not allele-specific, inhibition of transcripts encoding a truncated version of the human androgen receptor with different trinucleotide repeat lengths was observed in *Drosophila* cells. In human cells, small dsRNAs specifically inhibited similar truncated *ar* containing transcripts and further this inhibition protected cells from the cytotoxic effects associated with the accumulation of truncated AR protein containing an expanded polyQ tract.

In invertebrates such as *C. elegans* and *Drosophila*, dsRNA-mediated RNAi is most effectively triggered by RNA duplexes of over about 80 nts. However, several recent studies have been shown that the larger input dsRNA is cleaved into fragment ~21-25 nts that act as guides for the enzymatic complex required for the degradation of the target mRNA (Hammond *et al.*, *Nature* 404:293-296, 2000; Parrish *et al.*, *Mol. Cell* 6:1077-1087, 2000; Zamore *et al.*, *Cell* 101:25-33, 2000; Yang *et al.*, *Curr. Biol.* 10:1191-1200, 2000; Bernstein *et al.*, *Nature* 409:363-366, 2001; Hamilton

and Baulcombe, *Science* 286:950-952, 1999; Elbashir *et al.*, *Genes. Dev.* 15:188-200, 2001). This observation may explain detection of gene specific (but not allele specific) inhibition, since cleavage of the *arcag26ar*, *arcag43ar*, or *arcag106ar* dsRNAs to dsRNAs of ~21-25 nts will likely result in a loss of any duplex that distinguishes these from one another. This has significance for the application for this methodology to dominant genetic disorders, where the only distinguishing feature is a trinucleotide expansion. In the case of X-linked SBMA, which is mono-allelic, this is potentially less of an issue though androgen insensitivity may be exacerbated, but where loss of function of the normal allele may have a broader clinical impact, this observation will need to be considered.

Though the *arcag$_n$ar* dsRNAs did not confer allele specificity, the addition of larger *ar* flanking sequences did protect against the non-specific effects seen when dsRNAs generated from oligoribonucleotides with 27 CAG repeats were used. There are several possible reasons for this; one is that the CAG repeat results in a complex secondary folding that generates a toxic duplex. However, no overt cytotoxicity and no effect on *gapdh-1* mRNA levels were observed.

Alternatively, there is the potential that the small dsRNAs generated from the *cag27* and *arcag27* dsRNAs can targeting a broader ranger of transcripts that contain a small repeat, resulting in decreased expression from the parental plasmid (pAct.GFP) as well as the transcripts we were aiming to target. However, in both cases the *cag27* and the *arcag27* dsRNAs did induce a greater inhibition of the *arcag$_n$ar* transcripts than the parental plasmid suggesting at least a level of targeted degradation was occurring.

Sequence-specific dsRNA mediated interference of gene expression has only recently been observed in mammalian cell culture systems (Elbashir *et al.*, *Nature* 411:494-498, 2001; Caplen *et al.*, *Proc. Natl. Acad. Sci. U. S. A.* 98:9742-9747, 2001). This example demonstrates that the same mechanism can be used to specifically target a transcript associated with a dominant genetic disorder and to rescue the adverse cellular changes associated with the expression of a transcript containing a trinucleotide expansion. Though the inhibition mediated in this study used a transient co-transfection system, there is evidence that the same small dsRNA methodology can be used to inhibit endogenous gene expression.

All three of the experimental 22 nt dsRNAs used in this example inhibited expression of the target transcript, though some variation was observed in the degree of rescue mediated by the two *ar* dsRNAs, with the *5'ar* dsRNA proving more effective that the *3'ar* dsRNA. Both of these dsRNAs have similar sequence complexity and structure. However, the *5'ar* dsRNA targeted a region overlapping the start of the *ar* open-reading frame about 160 nts from the start if the CAG tract whereas the *3'ar* dsRNA targeted a sequence approximately 90 nts downstream of the CAG tract and it may be these features of the target sequence that explain the slight variations we observed.

This example demonstrates the feasibility of targeting a disease-associated transcript by dsRNA mediated RNAi. In this example dsRNAs against the androgen receptor in the context of the genetic disease SBMA, however, knockdown of the androgen receptor may also have clinical application in the treatment of cancers, in particular prostate cancer.

Example 5: High-throughput analysis of knock-downs (inhibition/interference) by small dsRNAs identifies potential clinical targets and lead drug compounds

One of the most important and well-characterized oncogenes in human cancer is the estrogen receptor tyrosine kinase Erb-B2, or HER2. Erb-B2 is expressed in normal cells at very low levels and is a very powerful transducer of growth signals from hormone ligands. It is significantly over-expressed in approximately 30% of breast cancers and a significant number of ovarian cancers. Erb-B2 overexpression is also a marker for a more aggressive tumor and poor prognosis. Erb-B2 was one of the first protein targets for which a specifically designed molecular-based therapy was developed. Herceptin is an antibody that specifically targets the Erb-B2 oncogene and sterically inhibits ligand binding and thus the subsequent activation of downstream pathways that result in uncontrolled proliferation of cells. This steric inhibition by Herceptin thus results in decreased tumor cell growth and invasion. Approximately 10% of breast tumors over-expressing Erb-B2 respond to this monoclonal antibody therapy. Though this antibody based means of treatment represents a significant therapeutic step forward, still a large number of Erb-B2 positive tumors are resistant to Herceptin and thus alternative therapeutics are still required.

Inhibition of Erb-B2 gene expression at the mRNA level using small dsRNA RNAi opens up a new realm of therapeutic potential for breast cancer and many other types of human malignancies. In this example a small dsRNA against the human *Erb-B2* sequence was assessed to determine the effect of down–regulating this oncogene in a breast cancer cell line (SKBR3). Further, we examine in a high-throughput manner a series of small dsRNAs against *Erb-B2* to see if a small dsRNA that could act as a lead drug compound could be identified.

Methods

Nucleic acids. Single stranded, gene specific sense and antisense oligoribonucleotides were synthesized using 2'-O-(tri-isopropyl) silyloxymethyl chemistry by Xeragon Inc. (Huntsville, AL; Gaithersburg, MD). Double stranded RNA molecules were generated by mixing sense and antisense ssRNA oligomers (50-200 μg each) in 10 mM Tris-Cl (pH 7.0), 20 mM NaCl (total volume 300 μl), heating to 95°C, and by incubating for 1 hr at 37°C or by cooling slowly (18 hours) to room temperature. The integrity and the dsRNA character of the annealed RNAs were confirmed by gel electrophoresis. Sequences of small dsRNAs used in this Example 5 are shown in Table 6.

Table 6: Human ERB-B2 dsRNA Sequences

SEQ ID NO:	Name	Sequence	Residues
178	*ERBOS*	aagugugcaccggcacagacadTdT	
179	*ERBOAS*	ugucugugccggugcacacTdT	
47	*ERB1S.*	r(aggcacccagcucuuugaggac)	460-481
48	*ERB1A.*	r(ccucaaagagcugggugccucg)	458-479
49	*ERB2S.*	r(aacuaugcccuggccgugcuag)	482-503
50	*ERB2A.*	r(agcacggccagggcauaguugu)	480-501
51	*ERB3S.*	r(acaauggagacccgcugaacaa)	504-525

52	*ERB3A*	r(guucagcgggucuccauugucu)	502-523
53	*ERB4S.*	r(uaccaccccugucacaggggcc)	526-547
54	*ERB4A.*	r(ccccugugacaggggugguauu)	524-545
55	*ERB5S.*	r(uccccaggaggccugcgggagc)	548-569
56	*ERB5A.*	r(ucccgcaggccuccuggggagg)	546-567
57	*ERB6S.*	r(ugcagcuucgaagccucacaga)	570-591
58	*ERB6A.*	r(ugugaggcuucgaagcugcagc)	568-589
59	*ERB7S.*	r(gaucuugaaaggaggggucuug)	592-613
60	*ERB7A.*	r(agaccccuccuuucaagaucuc)	590-611
61	*ERB8S.*	r(auccagcggaacccccagcucu)	614-635
62	*ERB8A.*	r(agcuggggguuccgcuggauca)	612-633
63	*ERB9S.*	r(gcuaccaggacacgauuuugug)	636-657
64	*ERB9A.*	r(caaaaucguguccugguagcag)	634-655
65	*ERB10S*	r(gaacaauaccaccccugucaca)	520-541
66	*ERB10A*	r(ugacaggggugguauuguucag)	518-539
67	*ERB11S*	r(uggggagagaguucugaggauu)	764-785
68	*ERB11A*	r(uccucagaacucucucccagc)	762-783
69	*ERB12S*	r(ccagugccaauauccaggaguu)	1221-1242
70	*ERB12A*	r(cuccuggauauuggcacuggua)	1219-1240
71	*ERB13S*	r(aguaauccggggacgaauucug)	1438-1459
72	*ERB13A*	r(gaauucgucccggauuacuug)	1436-1457
73	*ERB14S*	r(uccaguggccaucaaaguguug)	2395-2416
74	*ERB14A*	r(acacuuugauggccacuggaau)	2393-2414
75	*ERB15S*	r(agacgaagcauacgugauggcu)	2455-2476
76	*ERB15A*	r(ccaucacguaugcuucgucuaa)	2453-2474
77	*ERB16S*	r(acugugugggagcugaugacuu)	2882-2903
78	*ERB16A*	r(gucaucagcucccacacaguca)	2880-2901
79	*ERB17S*	r(cuggcuccgauguauuugaugg)	3393-3414
80	*ERB17A*	r(aucaaauacaucggagccagcc)	3391-3412
81	*ERB18S*	r(uggggucgucaaagacguuu)	3685-3704
82	*ERB18A*	r(acgucuuugacgaccccauu)	3683-3702
83	*ERB19S*	r(ccuguccuaaggaaccuuccuu)	4048-4069
84	*ERB19A*	r(ggaagguuccuuaggacagguu)	4046-4067
85	*ERB20S*	r(uggugucaguauccaggcuu)	4356-4375
86	*ERB20A*	r(gccuggauacugacaccauu)	4354-4373
87	*ERB21S*	r(gugugcaccggcacagacaug)	222-243
88	*ERB21A*	r(ugucugugccggugcacacuu)	220-241
89	*ERB22S*	r(gttaccagtgccaatatccag)	1216-1236
90	*ERB22A*	r(ggatattggcactggtaactg)	1214-1234
91	*ERB23S*	r(agtacacgatgcggagactgc)	2201-2221
92	*ERB23A*	r(agtctccgcatcgtgtacttc)	2199-2219
93	*ERB24S*	r(attgactctgaatgtcggcca)	3031-3051
94	*ERB24A*	r(gccgacattcagagtcaatca)	3029-3049
95	*ERB25S*	r(tacagcggtacagtgaggacc)	3476-3496
96	*ERB25A*	r(tcctcactgtaccgctgtaga)	3474-3494

Cell culture, nucleic acid transfections and analysis of gene expression. SKBR3 cells (a human breast cancer cell line), American Type Culture Collection (ATCC), Manassas, VA) were grown in McCoys Medium (Gibco, Life Technologies, Gaithersburg, MD) supplemented with 10% FBS (X). RNA (0.5-2 μg) was complexed with Lipofectamine at a nucleic acid: cationic lipid ratio of 1:5 in OptiMEM for ~15 minutes. Three to four hours after initiation of transfection, McCoys medium supplemented with 20% FBS was added to cells. Erb-B2 expression was assessed in

mammalian cells by Western analysis using an Erb-B2 specific antibody, fluorescence activated cell analysis (FACs) (FacsCaliber, Becton Dickinson, San Jose, CA) using an anti-erbb2(p185)-FITC(Novocastra) antibody and untransfected cells to control for background fluorescence. For analysis of cells transfected in 8 well chamber slides at 72 hours post transfection, slides were washed in 1x Dulbelco's phosphate buffered saline (DPBS) and fixed with 2% Paraformaldehyde. Cells were permeabilized with cold 0.2% triton-x in 1xDBPS and stored in fresh DPBS until the assays were performed. Cell proliferation rates were assayed using an antibody against the proliferation marker ki-67 (anti-ki67-FITC, Novocastra, UK). Apoptotic index was measured using the Tunel DNA fragmentation assay (Clontech Inc. Palo Alto, USA) and both assays use a DAPI nuclear counterstain. Monitoring expression of Erb-B2 was done by immunfluorescent staining with an anti-her2(p185)-FITC antibody (Novacastra, UK). Images of immunofluorescently stained cells were taken using an AxioPlan2 microscope fitted with a Marchauser 8 position stage and an Axiocam CCD digital camera (Carl Zeiss GmbH, Oberkochen, Germany) at 10x magnification. Automated image capture was done using custom macro scripting in the KS400 software package (Carl Zeiss GmbH, Oberkochen, Germany).

To assess the relative effectiveness of each small dsRNA, graphs were plotted that represent the percentage of the cell population in the given image at each of the 256 pixel intensity thresholds. This graphical representation is essentially a summary of the percentage of positive cells for any possible chosen threshold. These graphs have richer information than a simple number of the percentage of positive cells for a single threshold, which is the popular characterization used to estimate the percent of positive cells in the image (see Example 5 Fig. 24, and Example 6 Figs. 26 and 27).

Results and Discussion

A small dsRNA corresponding to the erb-b2 gene can significantly reduce the levels of Erb-B2 protein and can lead to a reduction in cell proliferation.

To demonstrate that a small dsRNA against Erb-b2 can down-regulate the Erb-B2 protein SKBR3 cells were transfected with a small dsRNA *ERBO* using a cationic lipid as carrier and protein lysates harvested from the transfected cells 72 hours later. Using standard Western blot analysis applying antibodies against ErB2, cyclin D and a control protein (tubulin) a significant decrease in the levels of Erb-B2 protein was seen. Importantly the levels of a protein that is dependent on the expression of Erb-B2, cyclin D, were also significantly decreased (Fig. 22A).

We also assessed this specific decrease in Erb-B2 expression mediated by the *ERBO* small dsRNA by using FACs analysis of SKBR3 transfected cells (Fig 22B). At 48 hours after transfection a significant shift of the fluorescent signal corresponding to Erb-B2 was seen. The decrease in Erb-b2 expression following transfection of the *ERB0* small dsRNA could also be visualized using immunohistochemistry (Fig. 23A). The decrease Erb-b2 expression was also associated with a decrease in the expression of the Ki67 antigen, which acts as a marker of proliferation (Fig 23B). These experiments show that a gene strongly associated with a large number of malignancies can be

targeted by a small dsRNA and that the down-regulation of this genes results in a phenotype consistent with a decrease in the proliferation and survival of these cells suggesting that this approach could be used as therapeutic strategy.

Screening of multiple small dsRNAs against *Erb-B2* as a means of identifying lead molecules that can be used as a potential thereaputic.

An important feature of the adaptation of small dsRNA mediated RNAi in mammalian cells as a technology is that, in a variety of platforms including the RNAi microarray system described in Example 3 and the high throughput screening of cells transfected in chamber slides or in plastic wells including, though not limited to, 96 wells, it is possible to rapidly screen for novel target genes and optimized small dsRNAs that can be used for a variety of therapeutic strategies including the inhibition of tumorgenesis.

To assess the feasibility of this approach, we screened 25 different dsRNAs against Erb-B2 in SKBR3 cells using the image quantifications system described in Example 3 and above. SKBR3 cells were grown and transfected in 8-well chamber slides duplicate wells were transfected with 25 different Erb-B2 small dsRNAs (Table 6). After 72 hours the slides were fixed, subjected to immunohistochemistry for Erb-b2 and the level of Erb-B2 expression quantified by a script which picks three random positions inside each well of the 8 well chamber slides, autofocuses in each channel, takes an image in each color, overlayes the colors and normalizes the output image to give a true color overlay, and stores the files with a unique filename identifier in the database. Image analysis was performed by custom algorithms written in Matlab6.1 that scan directories, match up image sets (Red,Green,Blue, and RGB overlay) and then segments and quantitates the image on a cell-by-cell basis, correcting for local background around each cell perimeter.

Data output was a final nuclear mask showing the exact nuclear segmentation of the cells, a final cytoplasmic mask showing the segmentation of the full cells, a curve showing the percentage of cells above any given threshold mean intensity, statistical analysis of every cell in an image (including intensity minimum, maximum, mean, median, standard deviation, nuclear area, cytoplasmic area, and total cell area), and a table showing mean image values summarizing the data in an entire image; over 2000 images were analyzed in this way. Figure 24 illustrates Erb-B2 expression in SKBR3 cells transfected with various small dsRNAs, subjected to immunohistochemistry for Erb-B2 expression. The graph represents the percentage of Erb-B2 positive cells at any mean threshold intensity. In this graph the data for four different *Erb-B2* small dsRNAs (ERB9, SEQ ID NOs: 63 and 64, Erb10, SEQ ID NOs: 65 and 66, ERB 11, SEQ ID NOs: 67 and 68, ERB14, SEQ ID NOs: 73 and 74) is shown all of which show a significant decrease in Erb-B2 expression relative to the negative control (untransfected or *egfp* small dsRNA transfected cells SEQ ID NOs: 3 and 4). To further illustrate the utility of this approach Table 7 shows the percentage of Erb-B2 positive cells seen in cells transfected with the stated *Erb-b2* small dsRNA

Table 7

Slide Code	Small dsRNA	% positive
HER2#10_1CA	Negative control	80.4
HER2#10_1CC	*egfp*	65
HER2#10_1CD	*egfp*	71.9
HER2#11_10CA	Negative control	78.9
HER2#11_10CB	Negative control	80.7
HER2#11_10CC	*egfp*	81.3
HER2#11_10CD	*egfp*	84.2
HER2_#11_1C_A	Negative control	73.6
HER2_#11_1C_B	Negative control	71.1
HER2_#11_1C_C	*egfp*	86.7
HER2_#11_1C_D	*egfp*	80.5
HER2#10_2CA	*ERB1*	59.2
HER2#10_2CB	*ERB1*	70.6
HER2#10_2CC	*ERB2*	72.4
HER2#10_2CD	*ERB2*	63.5
HER2#10_2CE	*ERB3*	67.6
HER2#10_2CF	*ERB3*	65.1
HER2#10_2CG	*ERB4*	65.2
HER2#10_2CH	*ERB4*	70.4
HER2#10_4C_A	*ERB9*	50.2
HER2#10_4C_B	*ERB9*	50.2
HER2#10_4C_C	*ERB10*	32.9
HER2#10_4C_D	*ERB10*	40.8
HER2#10_4C_E	*ERB11*	51
HER2#10_4C_F	*ERB11*	41.8

Slide Code	Small dsRNA	% positive
HER2#10_4C_G	*ERB12*	55.9
HER2#10_4C_H	*ERB12*	40
HER2#10_5C_A	*ERB13*	51.3
HER2#10_5C_B	*ERB13*	22.5
HER2#10_5C_C	*ERB14*	42.8
HER2#10_5C_D	*ERB14*	32.1
HER2#10_5C_E	*ERB15*	38.4
HER2#10_5C_F	*ERB15*	32
HER2#10_5C_G	*ERB16*	30.7
HER2#10_5C_H	*ERB16*	35.7
HER2_#10_3C_A	*ERB5*	65.8
HER2_#10_3C_B	*ERB5*	40.1
HER2_#10_3C_C	*ERB6*	55.2
HER2_#10_3C_D	*ERB6*	48
HER2_#10_3C_F	*ERB7*	52.7
HER2_#10_3C_H	*ERB8*	58.2
HER2_#10_7C_A	*ERB22*	81.1
HER2_#10_7C_B	*ERB22*	84.7
HER2_#10_7C_C	*ERB23*	43.5
HER2_#10_7C_D	*ERB23*	34.2
HER2_#10_7C_E	*ERB24*	31.7
HER2_#10_7C_F	*ERB24*	16.6
HER2_#10_7C_G	*ERB25*	36.3
HER2_#10_7C_H	*ERB25*	50.3

To further validate this type of screening approach, we confirmed that more than one of the *Erb-B2* small dsRNAs that showed an ability to mediate RNAi against *Erb-B2* could also induce apoptosis by transfecting SKBR3 cells with either a control small dsRNA (egfp) or with ERB0 or ERB5. Figure 25 shows that both of these small dsRNAs do induce a significantly higher level of apoptosis than a control small dsRNA

Example 6. High throughput screening to identify/validate anti-cancer targets and therapeutic small dsRNAs.

In Examples 3 and 5 it was established that multiple dsRNAs against the same gene target can be assessed in rapid, high through-put manner for their phenotypic effect. In this example it is established that multiple small dsRNAs against multiple gene targets can be assessed using similar approaches and that in doing so novel anti-cancer targets and therapeutic small dsRNAs can be identified.

Methods

Single stranded, gene specific sense and antisense oligoribonucleotides were synthesized and dsRNA molecules formed essentially as described in Examples 1-5. The integrity and the

dsRNA character of the annealed RNAs were confirmed by gel electrophoresis. Two or three small dsRNAs per gene target were designed for 31 genes associated with the development or progression of a variety of cancers including breast and prostate cancer (Table 8).

Table 8

SEQ ID NO:	Name	Sequence (5' to 3')	nt position	Length
180	APPBP2 S1	GUGGACAACUACAUCCGCUCC		21
181	APPBP2 A1	AGCGGAUGUAGUUGUCCACGA	370	21
182	APPBP2 S2	GAGGUUGCUUCAUGUGCGAAA		21
183	APPBP2 A2	UCGCACAUGAAGCAACCUCAC	807	21
184	BMP7 S1	CUACAAGGACUACAUCCGGGA		21
185	BMP7 A1	CCGGAUGUAGUCCUUGUAGAU	668	21
186	BMP7 S2	UUCGAUGACAGCUCCAACGUC		21
187	BMP7 A2	CGUUGGAGCUGUCAUCGAAGU	1365	21
188	BCLXL S1	UGAACUCUUCCGGGAUGGGdTdT		
189	BCL-XL AS1	CCCAUCCCGGAAGAGUUCAdTdT		
190	BCL-XL S2	GGAGAUGCAGGUAUUGGUGdTdT		
191	BCL-XL AS2	CACCAAUACCUGCAUCUCCdTdT		
192	BCL-XL S3	CUCUUCCGGGAUGGGGUAAdTdT		
193	BCL-XL AS3	UUACCCCAUCCCGGAAGAGdTdT		
194	CCND1 S1	GAUCGUCGCCACCUGGAUGCU		21
195	CCND1 A1	CAUCCAGGUGGCGACGAUCUU	401	21
196	CCND1 S2	UGCAUGUUCGUGGCCUCUAAG		21
197	CCND1 A2	UAGAGGCCACGAACAUGCAAG	543	21
198	CCND1 S3	CUCCUACGAUACGCUACUAUA		21
199	CCND1 A3	UAGUAGCGUAUCGUAGGAGUG	1926	21
200	CRYM S1	UUGGUCACCUUCUACGAGGAC		21
201	CRYM A1	CCUCGUAGAAGGUGACCAACU	274	21
202	CRYM S2	AUGAUUCCUGGUCAUCUGGUA		21
203	CRYM A2	CCAGAUGACCAGGAAUCAUAG	968	21
204	ER1 S1	GACUAUGCUUCAGGCUACCAU		21
205	ER1 A1	GGUAGCCUGAAGCAUAGUCAU	946	21
206	ER1 S2	AUGAUUGGUCUCGUCUGGCGC		21
207	ER1 A2	GCCAGACGAGACCAAUCAUCA	1540	21
208	EGFR S1	GAGUAACAAGCUCACGCAGUU		21
209	EGFR A1	CUGCGUGAGCUUGUUACUCGU	306	21
210	EGFR S2	UCAUACGCGGCAGGACCAAGC		21
211	EGFR A2	UUGGUCCUGCCGCGUAUGAUU	1478	21
212	EGFR S3	AUGAUAGACGCAGAUAGUCGC		21
213	EGFR A3	GACUAUCUGCGUCUAUCAUCC	3058	21
214	FKBP5 S1	GAGAAGACCACGACAUUCCAA		21
215	FKBP5 A1	GGAAUGUCGUGGUCUUCUCCU	749	21
216	FKBP5 S2	UGAACCUGGCCAUGUGCUACC		21
217	FKBP5 A2	UAGCACAUGGCCAGGUUCAGA	1133	21
218	FKBP5 S3	GACACGCAGUUGCUCGCUUUU		21
219	FKBP5 A3	AAGCGAGCAACUGCGUGUCAA	2734	21
220	FLJ20940 S1	GAGCUGAUCUUAGAUCCGGCG		21

221	FLJ20940 A1	CCGGAUCUAAGAUCAGCUCAA	214	21
222	FLJ20940 S2	GACUAAGGCUGGUCUGUGUUU		21
223	FLJ20940 A2	ACACAGACCAGCCUUAGUCUU	889	21
224	GRB7 S1	GAUGUGAACGAGUCCAACGUG		21
225	GRB7 A1	CGUUGGACUCGUUCACAUCUG	1066	21
226	GRB7 S2	CCUUCCGCCUCUUCAAGUACG		21
227	GRB7 A2	UACUUGAAGAGGCGGAAGGCA	1232	21
228	HOXB7 S1	UCCAGCCUCAAGUUCGGUUUU		21
229	HOXB7 A1	AACCGAACUUGAGGCUGGAUA	159	21
230	HOXB7 S2	AACUGACCGCAAACGAGGCCG		21
231	HOXB7 A2	UAGAUCCGGAAGUUACUCUCG	479	21
232	LMO4 S1	UUCCUGCGAGUGAACUCGUCA		21
233	LMO4 A1	ACGAGUUCACUCGCAGGAAUC	1079	21
234	LMO4 S2	CAGAGUAAUGCAGAAUGCGUG		21
235	LMO4 A2	CGCAUUCUGCAUUACUCUGAC	1302	21
236	MGC9753 S1	AUCCGCUUCGACUAUGGCUAC		21
237	MGC9753 A1	AGCCAUAGUCGAAGCGGAUGA	684	21
238	MGC9753 S2	ACUGAGUGUGCUCUUAGCUCU		21
239	MGC9753 A2	AGCUAAGAGCACACUCAGUUC	1257	21
240	MLN64 S1	UCUUCCGCUUCUCUGGACUGC		21
241	MLN64 A1	AGUCCAGAGAAGCGGAAGAAG	447	21
242	MLN64 S2	CUCCUAUGACGUGUCUGCAGG		21
243	MLN64 A2	UGCAGACACGUCAUAGGAGAU	1126	21
244	MYBL2 S1	AACCGCACUGACCAGCAAUGC		21
245	MYBL2 A1	AUUGCUGGUCAGUGCGGUUAG	332	21
246	MYBL2 S2	CUCAUCAUCGAGGACGACAUC		21
247	MYBL2 A2	UGUCGUCCUCGAUGAUGAGUU	1814	21
248	MYC S1	CAUCCAGGACUGUAUGUGGAG		21
249	MYC A1	CCACAUACAGUCCUGGAUGAU	963	21
250	MYC S2	AACUCUUGUGCGUAAGGAAAA		21
251	MYC A2	UUCCUUACGCACAAGAGUUCC	1882	21
252	NBS1 S1	ACUGGCGUUGAGUACGUUGUU		21
253	NBS1 A1	CAACGUACUCAACGCCAGUCA	182	21
254	NBS1 S2	CACAAGGCGUGUCAGUUGAUG		21
255	NBS1 A2	UCAACUGACACGCCUUGUGAA	1098	21
256	NCOA3 S1	AUAUAACCGAGCAGUGUCUCU		21
257	NCOA3 A1	AGACACUGCUCGGUUAUAUGG	2784	21
258	NCOA3 S2	AAGCUCCUCCGCAACAGUUUC		21
259	NCOA3 A2	AACUGUUGCGGAGGAGCUUGU	4073	21
260	PIP5K2B S1	CCAGGAACGUGUUCAGCCAUC		21
261	PIP5K2B A1	UGGCUGAACACGUUCCUGGUA	1100	21
262	PIP5K2B S2	UCUGUUGACGUCUAUGCCAUG		21
263	PIP5K2B A2	UGGCAUAGACGUCAACAGAGG	1531	21
264	PNMT S1	ACCUCCGCAACAACUACGCGC		21
265	PNMT A1	GCGUAGUUGUUGCGGAGGUAG	145	21
266	PNMT S2	CCUCCGCACCUAUAUCAUGCC		21
267	PNMT A2	CAUGAUAUAGGUGCGGAGGUC	797	21
268	PPARBP S1	AAGGAUGCUCCUCUUCCAGAU		21
269	PPARBP A1	CUGGAAGAGGAGCAUCCUUGU	1340	21

270	PPARBP S2	CUUCCAGCCAUCCUAUGUCCU		21
271	PPARBP A2	GACAUAGGAUGGCUGGAAGAA	4212	21
272	PPM1D S1	AGUACAUGGAGGACGUUACUC		21
273	PPM1D A1	GUAACGUCCUCCAUGUACUUC	277	21
274	PPM1D S2	ACCAUGCGACGCAGACUUAGG		21
275	PPM1D A2	UAAGUCUGCGUCGCAUGGUGA	1956	21
276	RAD51C S1	ACUGCUGAGGAACUCCUAGAG		21
277	RAD51C A1	CUAGGAGUUCCUCAGCAGUCU	160	21
278	RAD51C S2	GUAGACCUUGCUACUGCCUGC		21
279	RAD51C A2	AGGCAGUAGCAAGGUCUACCA	568	21
280	RAE1 S1	AGAUCGCACAGCAUGAUGCUC		21
281	RAE1 A1	GCAUCAUGCUGUGCGAUCUGU	552	21
282	RAE1 S2	UACGCUUCCAGCUACGACUGG		21
283	RAE1 A2	AGUCGUAGCUGGAAGCGUAUG	1181	21
284	RPS6K S1	GACAUAGACCUGGACCAGCCA		21
285	RPS6K A1	GCUGGUCCAGGUCUAUGUCAA	130	21
286	RPS6K S2	AGCACCUGCGUAUGAAUCUAU		21
287	RPS6K A2	AGAUUCAUACGCAGGUGCUCU	1601	21
288	S100P S1	CAUGACGGAACUAGAGACAGC		21
289	S100P A1	UGUCUCUAGUUCCGUCAUGGU	39	21
290	S100P S2	GGACCUGGACGCCAAUGGAGA		21
291	S100P A2	UCCAUUGGCGUCCAGGUCCUU	216	21
292	TBX2 S1	UGGACAUUGUAGCCGCUGACG		21
293	TBX2 A1	UCAGCGGCUACAAUGUCCAUC	501	21
294	TBX2 S2	CUCUACGCUUGUACGAGGAGC		21
295	TBX2 A2	UCCUCGUACAAGCGUAGAGAC	945	21
296	TMEPAI S1	CUAUCCGUACCUGCAGCACGA		21
297	TMEPAI A1	GUGCUGCAGGUACGGAUAGGU	416	21
298	TMEPAI S2	GAACACUCCGCGCUUCUUAGA		21
299	TMEPAI A2	UAAGAAGCGCGGAGUGUUCUG	917	21
300	TRIM37 S1	ACAGAGCGUGGAGAGCAUUGC		21
301	TRIM37 A1	AAUGCUCUCCACGCUCUGUUC	72	21
302	TRIM37 S2	GUAGCGACAGUGACAUUGAAU		21
303	TRIM37 A2	UCAAUGUCACUGUCGCUACUC	2765	21
304	TXNIP S1	CAGACUUCGGAGUACCUGCGC		21
305	TXNIP A1	GCAGGUACUCCGAAGUCUGUU	432	21
306	TXNIP S2	UAAUUGGCAGCAGAUCAGGUC		21
307	TXNIP A2	CCUGAUCUGCUGCCAAUUACC	1123	21
308	ZNF217 S1	GUUCCAUUCCGAGCUACACAA		21
309	ZNF217 A1	GUGUAGCUCGGAAUGGAACAA	425	21
310	ZNF217 S2	UGUUACCGCAGGACUGUGUGU		21
311	ZNF217 A2	ACACAGUCCUGCGGUAACAGU	3150	21

To assess the effect of these small dsRNAs on such phenotypic endpoints as cell proliferation and apoptosis the breast cancer cell line (SKBR3) or the prostate cancer cell line (pc3) were grown in 8 well chamber slides and were transfected with dsRNA using the cationic lipid Lipofectamine (Life Technologies, Gaithersburg, MD) at a ratio of 1:5 RNA to lipid. Cells were

fixed in 2% paraformaldehyde 72 hours post-transfection and were analyzed essentially as described in Example 5.

Results and Discussion.

Quantitaive analysis of over 2000 images was used to assess the effect of transfecting small dsRNAs against approximately 30 genes associated with either breast or prostate cancer. Of the candidate genes previously shown to be genetically linked to breast cancer, small dsRNAS against three genes showed a clear effect on cell proliferation, RS6K (S6 kinase) and the 17q23 amplicon candidate genes TBX2, and GRB7 (Fig. 26), see also Table 8. Small dsRNAs against these three genes significantly decreased the rate of proliferation compared to a negative control. Many of the dsRNAs were also assessed for their ability to induce apoptosis of SKBR3 cells using a TUNEL assay for Apoptosis as a marker. Table 9 shows the apoptotic activity (fraction of Tunel-positive cells) for control transfected cells and SKBR3 cells transfected with small dsRNAs against either GRB7 or flj20940 (DARPP-32)

Table 9.

Small dsRNA	Apoptotic activity (fraction of Tunel-positive cells)
egfp-Rh (control)	0.01
egfp (control)	0.051
GRB7.1	0.008
GRB7.2	0.011
GRB7.2	0.001
FLJ20940.1	0.188
FLJ20940.2	0.542
FLJ20940.2	0.136

Both the small dsRNAs against the flj20940 gene appear to induce apoptosis in the SKBR-3 breast cancer cell line. Interestingly, the flj20940 gene has been identified as the DARPP-32 gene, a phosphatase inhibitor that is amplified and over-expressed in gastric cancers (El-Rifai *et al.*, *Cancer Res* 2002, 15:4061).

Of the genes studied that may play a role in the development of prostate cancer a screen of multiple small dsRNAs those against BXL-XL showed the most significant results (Fig. 27). BCL-XL is part of the BCL2 like family of proteins that regulate mitochondrial dependent apoptosis. Pellangaris *et al* (*Cell*, 109:321, 2002) illustrated that BCL-XL by itself could halt apoptotic signals transduced by c-myc. BCL-XL can inhibit the permeability transistion pore from forming thereby not allowing cytochrome c and AIF (apoptosis inducing factor) release from the mitochondria. It also can selectively inhibit apaf-1(WD40)/cytochrome c activation of pro-caspase 9 halting the caspase cascade. BCL-XL is the longer of 2 alternate splicings of the BCLX gene. BCL-XS is the shorter of the two sequences, missing one ~180bp exon. BCL-XS has exactly the opposite functions of BCL-XL; it is powerfully pro-apoptotic. By using small dsRNAs targeting only one of the splice variants(BCL-XL) it was hypothesized that this would alter the cellular ratio of BCL-XL to BCl-XS,

and possibly other BCL protein family members, so alleviating the apoptotic block inhibiting cell death.

Example 7: Gene silencing through RNA interference in the murine lung *in vivo*

This example provides a demonstration of the effectiveness of RNA interference in vivo, using a murine lung system.

Methods

The expression plasmid pCI-CAT carrying a chloramphenicol acetyl transferase reporter (*cat*) gene or an empty plasmid control (pCI) were complexed with cationic lipids (80 mg/mouse in a total volume of 100 ml). Double strand small dsRNAs corresponding to *cat* or *egfp* were complexed plasmid DNA and dsRNA were mixed prior to *in vivo* transfection. For lung transfection Balb/C mice (female, 6-8 weeks) were anaesthetised with an inhaled anaesthetic (metofane) and the liposome complexes were applied to the nose and "sniffed" voluntarily into the lung. Forty-eight hours after transfection the lungs were harvested, and homogenized. CAT expression was measured with a CAT ELISA according to manufacturers recommendations.

Results and Discussion

To establish if RNAi can be induced in a whole organism, a transient expression analysis of RNAi similar to that described in Example 1 was developed for the murine lung *in vivo*. The murine lung tissue can be very effectively transfected by nucleic acids complexed with a cationic lipid and instilled through the nose (Alton *et al., Nat Genet* 5:135, 1993, Wheeler *et al., Proc Natl. Acad. Sci. U. S. A.* 93:11454, 1996, .Scheule and Cheng *Adv. Drug Deliv. Rev.* 30:173, 1998). Using this system the effect of a small dsRNA against the marker gene chloramphenicol acetyl transferase (CAT) was assessed using a comparison of the co-transfection of a plasmid expressing CAT (pCI-CAT) and a control dsRNA (*egfp*) and a dsRNA against *cat*. The co-transfection of the *cat* dsRNA and pCI-CAT reduced CAT expression by approximately 90% compared with that seen in control animals (Fig. 26). As *in vitro* transient assays of RNAi correlate well with the knock-down of stable transgene and endogenous gene expression this example shows that RNAi can be induced in a whole mammal and thus small dsRNAs could be used *in vivo* to modulate gene expression as a means of studying gene function and as a therapeutic.

This disclosure provides methods of making dsRNAs for use in gene silencing, gene silencing arrays, and methods of their use. The disclosure further provides particular gene-silencing agents and methods of their use as pharmaceuticals. It will be apparent that the precise details of the compositions and methods described may be varied or modified without departing from the spirit of the described invention. We claim all such modifications and variations that fall within the scope and spirit of the claims below.

CLAIMS

What is claimed is:

1. An RNA for interference or inhibition of expression of a target gene, comprising double stranded RNA of about 15 to about 40 nucleotides in length and a 3' or 5' overhang having a length of 0-nucleotide to 5-nucleotides on each strand, wherein the sequence of the double stranded RNA is substantially identical to a portion of a mRNA or transcript of the target gene.

2. The RNA of claim 1, wherein the double stranded RNA contains about 19 to about 25 nucleotides.

3. The RNA of claim 2, wherein the double-stranded RNA contains 20 nucleotides.

4. The RNA of claim 2, wherein the double-stranded RNA contains 21 nucleotides.

5. The RNA of claim 2, wherein the double-stranded RNA contains 22 nucleotides.

6. The RNA of any one of claims 1 through 5, wherein the length of the 3' or 5' overhang is 0-nucleotide on at least one strand.

7. The RNA of claim 6, wherein the length of the 5' overhang is 0-nucleotide on both strands.

8. The RNA of any one of claims 1 through 5, wherein the length of the 3' or 5' overhang is 1-nucleotide to 5-nucleotides on at least one strand.

9. The RNA of claim 8, wherein the length of the 3' or 5' overhang is 2-nucleotides on at least one strand.

10. The RNA of claim 9, wherein the length of the 3' or 5' overhang is 2-nucleotides on both strands.

11. The RNA of claim 10, wherein the length of the 3' overhang is 2-nucleotides on both strands.

12. The RNA of claim 1, wherein the double-stranded RNA contains 20, 21, or 22 nucleotides, and wherein the length of the 3' overhang is 2-nucleotides on both strands.

13. The RNA of any one of claims 1 through 12, wherein the double-stranded RNA contains about 40-60% adenine+uracil (AU) and about 60-40% guanine+cytosine (GC).

14. The RNA of claim 13, wherein the double-stranded RNA contains about 50% AU and about 50% GC.

15. The RNA of any one of claims 1 through 14, further comprising at least one modified ribonucleotide.

16. The RNA of claim 15, wherein the modified ribonucleotide is in the sense strand of the double-stranded RNA.

17. The RNA of claim 15, wherein the modified ribonucleotide is in the 3' overhang of at least one strand.

18. The RNA of claim 17, wherein the modified nucleotide is in the 3' overhang of the sense strand.

19. The RNA of any one of claims 15 through 18, wherein the modified ribonucleotide comprises a detectable label, a thiophosphate nucleotide analog, a deoxynucleotide, a 2'-fluorouracil, a 2'-aminouracil, a 2'-aminocytidine, a 4-thiouracil, a 5-bromouracil, a 5-iodouracil, a 5-(3-aminoallyl)-uracil, an inosine, or a 2'O-Me-nucleotide analog.

20. The RNA of claim 19, wherein the modified ribonucleotide comprises a detectable label and the detectable label comprises a fluorophore.

21. The RNA of claim 20, wherein the fluorophore is rhodamine or FITC.

22. The RNA of claim 1, wherein the sequence of the double-stranded RNA is no more than 30% different from the portion of the mRNA or transcript of the target gene.

23. The RNA of claim 1, wherein the target gene is an endogenous gene in a cell.

24. The RNA of claim 23, wherein the cell is *in vivo* in an organism.

25. The RNA of claim 24, wherein the organism is an invertebrate animal or a vertebrate animal.

26. The RNA of claim 25, where the invertebrate animal is a nematode or a fly.

27. The RNA of claim 25, wherein the invertebrate animal is *C. elegans* or *Drosophila*

28. The RNA of claim 25, wherein the vertebrate animal is a mammal.

29. The RNA of claim 28, wherein the mammal is a mouse, a monkey, or a human.

30. The RNA of claim 1, comprising a sequence as shown in any one of SEQ ID NOs: 1 through 311.

31. A method of interfering with or inhibiting expression of a target gene in a cell, the method comprising exposing the cell to an effective amount of the RNA of any one of claims 1 through 23.

32. The method of claim 31, wherein the cell is an animal cell.

33. The method of claim 32, wherein the animal cell is a mammalian cell.

34. The method of claim 33, wherein the mammalian cell is a human cell.

35. The method of claim 32, wherein the animal cell is *in vivo*.

36. The method of claim 34, wherein the human cell is *in vivo*.

37. The method of claim 31, which is a genetic method for treating an animal.

38. The method of claim 31, further comprising repeating exposing the cell to the effective amount of the RNA of any one of claims 1 through 23.

39. A gene silencing array, comprising:

a substantially flat substrate; and

addressably arrayed thereon, a plurality of different double-stranded RNAs, each having a length of about 15 to 40 nucleotides, a 3' or 5' overhang having a length of 0-nucleotides to 5-nucleotides on each strand, and wherein the sequence of each double-stranded RNA is substantially identical to a portion of a mRNA or transcript of a target gene.

40. The gene silencing array of claim 39, wherein the sequence of at least 50% of the double-stranded RNAs is substantially identical to portions of a mRNA or transcript of a single target gene.

41. The gene silencing array of claim 40, wherein the sequence of at least 70% of the double-stranded RNAs is substantially identical to portions of a mRNA or transcript of a single target gene.

42. The gene silencing array of claim 40, wherein the single target gene is *Erb-B2*, *APPBP2*, *BMP7*, *CCND1*, *CRYM*, *ER1*, *FKBP5*, *FLJ20940*, *GRB7*, *HOXB7*, *LMO4*, *MGC9753*, *MLN64*, *MYBL2*, *MYC*, *NBS1*, *NCOA3*, *PIP5K2B*, *PNMT*, *PPARBP*, *PPM1D*, *RAD51C*, *RAE1*, *RPS6K*. *S100P*, *TBX2*, *TMEPAI*, *TRIM37*, *TXNIP*, or *ZNF217*.

43. The gene silencing array of any one of claims 39 through 42, comprising at least 50 different double stranded RNAs.

44. The gene silencing array of claim 43, comprising at least 100 different double-stranded RNAs.

45. The gene silencing array of claim 43, comprising at least 200 different double-stranded RNAs.

46. The gene silencing array of claim 43, comprising at least 500 different double-stranded RNAs.

47. The gene silencing array of claim 43, comprising at least 800 different double-stranded RNAs.

48. The gene silencing array of claim 43, comprising more than 1000 different double-stranded RNAs.

49. The gene silencing array of claim 39, wherein the substantially flat substrate is a microscope slide, a chamber slide, a culture plate, or a 96-well microtiter plate.

50. An array-based method of assessing a phenotypic effect of a double-stranded RNA on a target gene, comprising:

addressably arraying the double-stranded RNA on a substrate, wherein the double-stranded RNA has a length of about 15 to 40 nucleotides, a 3' or 5' overhang having a length of 0-nucleotides to 5-nucleotides on each strand, and wherein the sequence of the double-stranded RNA is substantially identical to a portion of a mRNA or transcript of the target gene;

overlaying the substrate with cells;

culturing the cells under conditions sufficient to enable at least a portion of the double-stranded RNA to be taken up into at least one of the cells; and

assessing at least one phenotypic characteristic of the at least one of the cells, wherein a change in the phenotypic characteristic in a cell that has taken up a double-stranded RNA compared to a cell that has not taken up the double-stranded RNA, or that has taken up a different double-stranded RNA, is indicative of a phenotypic effect.

51. The method of claim 50, wherein the cells are *in vitro* cultured animal cells.

52. The method of claim 51, wherein the *in vitro* cultured animal cells are mammalian cells.

53. The method of claim 52, wherein the mammalian cells are human cells.

54. The method of claim 53, wherein the animal cells are cancer cells.

55. The method of claim 50, wherein the target gene is *Erb-B2*, *APPBP2*, *BMP7*, *CCND1*, *CRYM*, *ER1*, *FKBP5*, *FLJ20940*, *GRB7*, *HOXB7*, *LMO4*, *MGC9753*, *MLN64*, *MYBL2*, *MYC*, *NBS1*, *NCOA3*, *PIP5K2B*, *PNMT*, *PPARBP*, *PPM1D*, *RAD51C*, *RAE1*, *RPS6K. S100P*, *TBX2*, *TMEPAI*, *TRIM37*, *TXNIP*, or *ZNF217*.

56. The method of claim 50, wherein the method is a method of assessing a phenotypic effect of a plurality of double-stranded RNAs on the target gene, further comprising addressably arraying the plurality of double-stranded RNAs on the substrate.

57. The method of claim 56, wherein the sequence of at least one of the plurality of double-stranded RNAs is substantially identical to a portion of a mRNA or transcript of a first target gene and the sequence of at least one of the plurality of the double-stranded RNAs is substantially identical to a portion of a mRNA or transcript of a second target gene, and wherein the method is a method of assessing a phenotypic effect of the plurality of double-stranded RNAs on a plurality of target genes.

58 The method of claim 57, wherein at least two of the plurality of target genes are selected from the group consisting of *Erb-B2*, *APPBP2*, *BMP7*, *CCND1*, *CRYM*, *ER1*, *FKBP5*, *FLJ20940*, *GRB7*, *HOXB7*, *LMO4*, *MGC9753*, *MLN64*, *MYBL2*, *MYC*, *NBS1*, *NCOA3*, *PIP5K2B*, *PNMT*, *PPARBP*, *PPM1D*, *RAD51C*, *RAE1*, *RPS6K. S100P*, *TBX2*, *TMEPAI*, *TRIM37*, *TXNIP*, and *ZNF217*.

59. The method of claim 50, wherein the phenotypic effect is an anti-cancer phenotypic effect.

60. The method of claim 59, wherein the anti-cancer phenotypic effect comprises an anti-proliferative effect, a pro-apoptotic effect or a combination thereof.

61. The method of claim 50, wherein the substrate is a microscope slide, a chamber slide, a culture plate, or a 96-well microtiter plate.

62. A method of validating a gene as a potential drug target for a disease or condition, comprising:

assaying a plurality of dsRNAs, each having a sequence substantially identical to a portion of a mRNA or transcript of the gene, for the ability to generate a desired phenotype in a cell, wherein the phenotype is a phenotype related to the disease or condition; and

selecting at least one dsRNA that generates the desired phenotype.

63. The method of claim 62, wherein the gene is a cancer-associated gene and the desired phenotype of the cell comprises an anti-cancer phenotype.

64. The method of claim 63, wherein the anti-cancer phenotype comprises inhibition of proliferation, stimulation of apoptosis, or a combination thereof.

65. The method of claim 62, wherein assaying the plurality of dsRNAs comprises an array-based analysis.

66. A method of selecting an optimized sequence of a double-stranded RNA for interference with or inhibition of expression of a target gene in a cell, comprising:

selecting one or more sub-sequence from within the sequence of the target gene that satisfy a first criterion, where the first criterion is that each sub-sequence is about 15 to about 40 nucleotides long;

selecting one or more sub-sequence from within the sequence of the target gene that satisfy a second criterion, where the second criterion is that each sub-sequence has a AU:GC content percent of about 40:60 to 60:40;

selecting one or more sub-sequence from within the sequence of the target gene that satisfy a third criterion, where the third criterion is that each sub-sequence is unique in comparison to a gene database from the same organism as the cell; and

selecting at least one sub-sequence that satisfies the first, second, and third criteria, which sub sequence is an optimized sequence of the double-stranded RNA for interference with or inhibition of expression of the target gene in the cell.

67. The method of claim 66, wherein the first criterion is that each sub-sequence is about 19 to 23 nucleotides long.

68. The method of claim 66, wherein the second criterion is that each sub-sequence has a AU:GC content percent of about 50:50.

69. The method of claim 66, wherein selecting one or more sub-sequences to satisfy the first, second, and third criteria occur concurrently.

70. The method of claim 66, wherein selecting one or more sub-sequences to satisfy the first, second, and third criteria occur sequentially.

71. A short double-stranded RNA effective for interfering with or inhibiting expression of a target gene, comprising a sequence as shown in any one of SEQ ID NOs: 1 through 311.

Figure 1

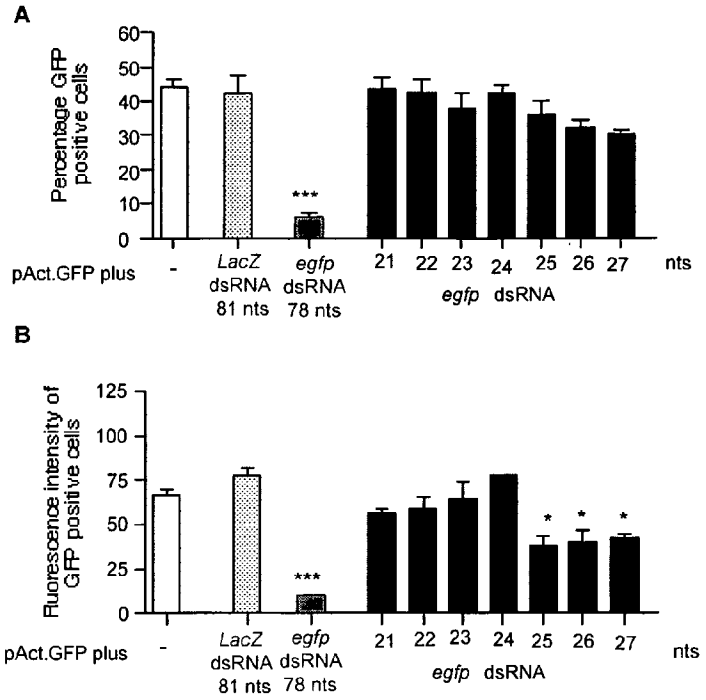


Figure 4

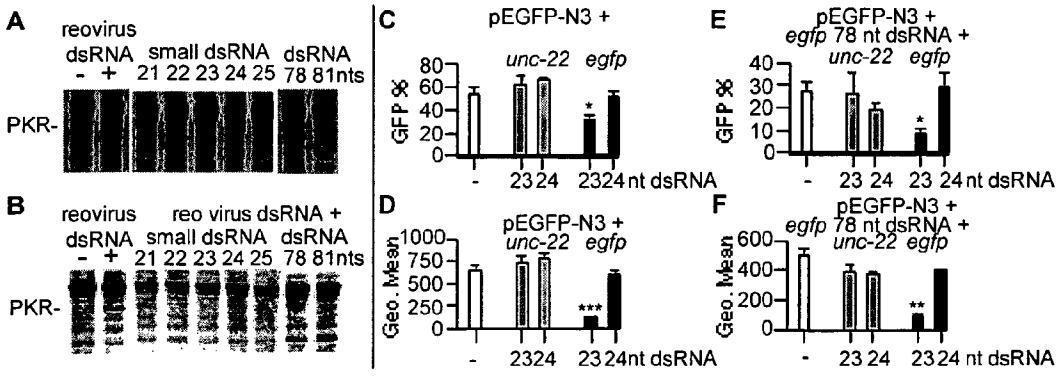


Figure 2

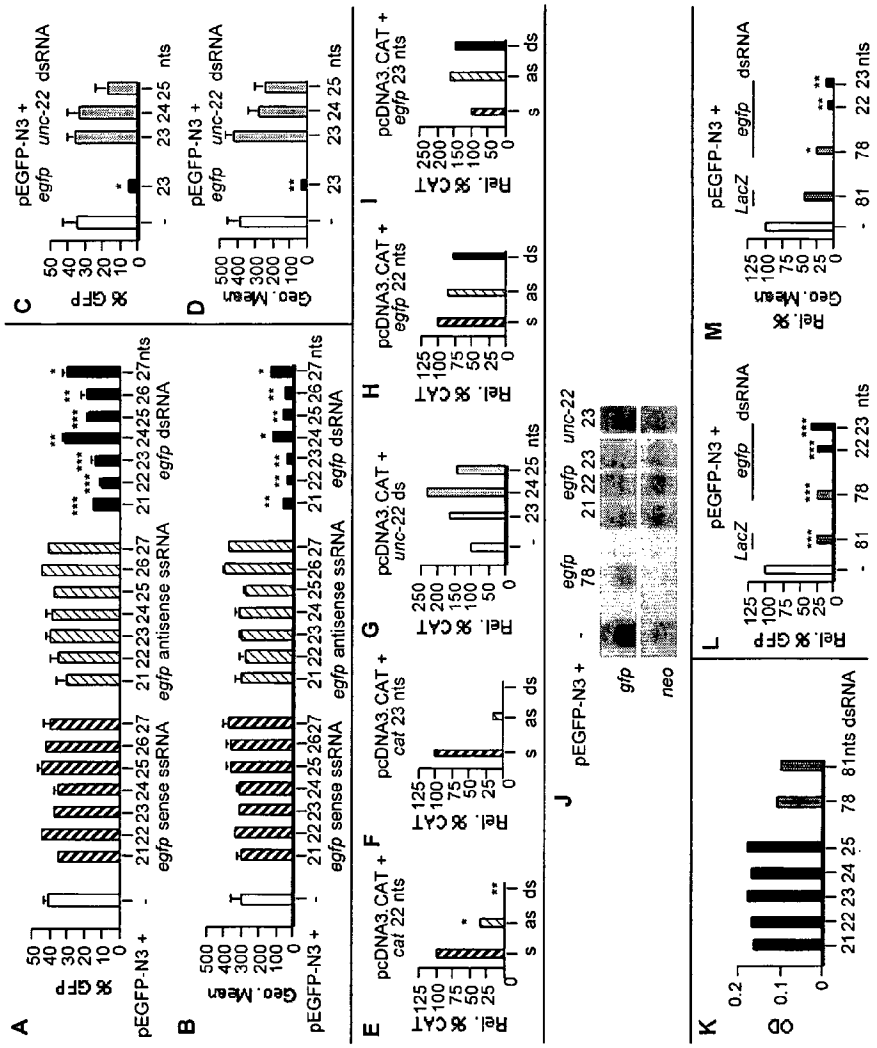

A
% GFP
pEGFP-N3 + -
21 22 23 24 25 26 27
egfp sense ssRNA
egfp antisense ssRNA
egfp dsRNA
27nts
B
Geo. Mean
pEGFP-N3 +
egfp sense ssRNA
egfp antisense ssRNA
egfp dsRNA
C
pEGFP-N3 +
egfp unc-22 dsRNA
- 23 23 24 25 nts
D
Geo. Mean
E
pcDNA3.CAT +
cat 22 nts
Rel. % CAT
s as ds
F
cat 23 nts
G
unc-22 ds
23 24 25 nts
H
egfp 22 nts
I
egfp 23 nts
J
pEGFP-N3 +
egfp
unc-22
gfp
neo
K
OD
21 22 23 24 25 78 81nts dsRNA
L
LacZ
egfp dsRNA
- 81 78 22 23 nts
M
LacZ
egfp dsRNA
- 81 78 22 23 nts

Figure 3

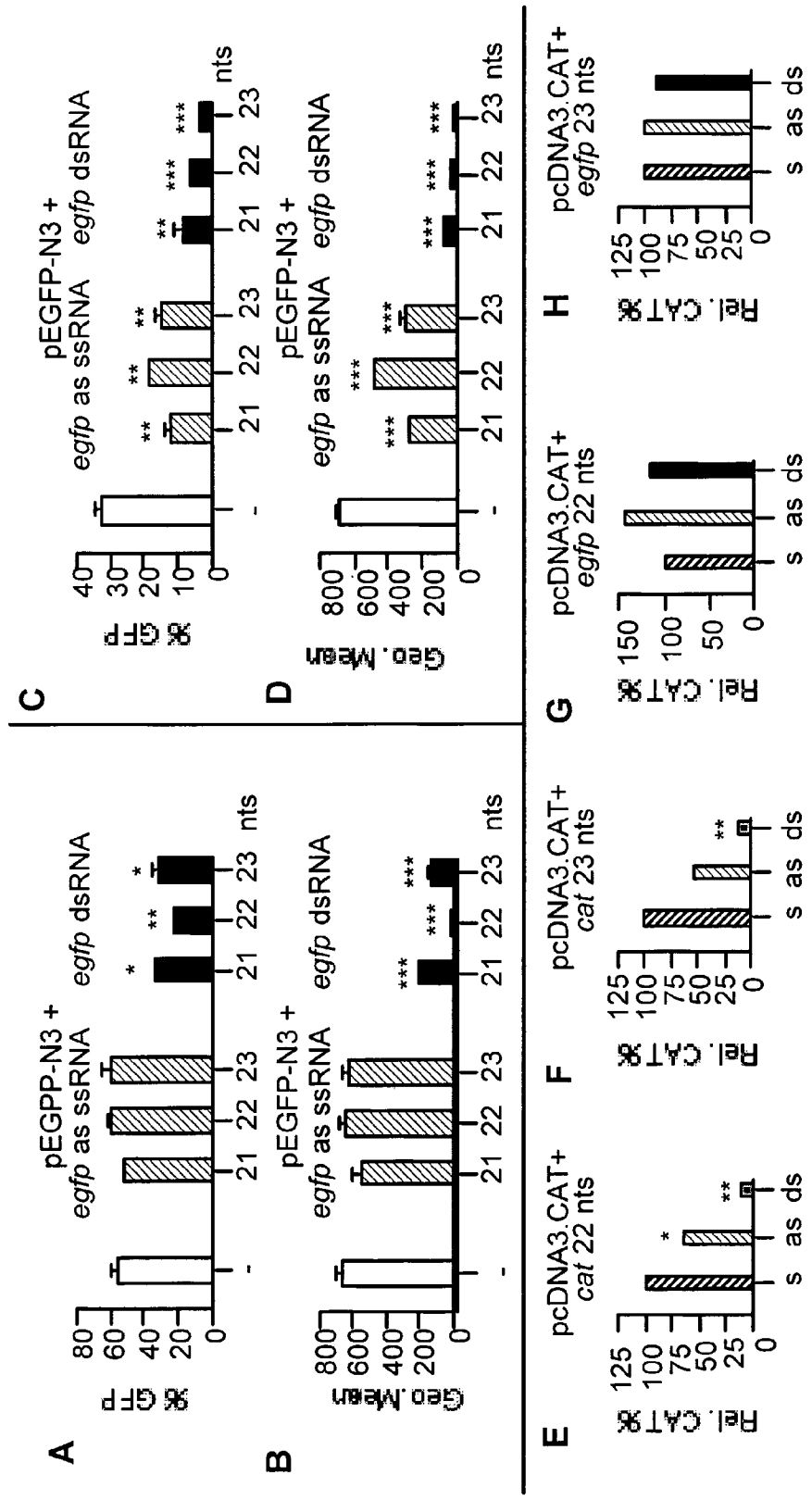

A
% GFP
pEGPP-N3 + egfp as ssRNA egfp dsRNA
- 21 22 23 21 22 23 nts
B
Geo.Mean
pEGFP-N3 + egfp as ssRNA egfp dsRNA
C
% GFP
pEGFP-N3 + egfp as ssRNA egfp dsRNA
D
Geo.Mean
pEGFP-N3 + egfp as ssRNA egfp dsRNA
E
Rel. CAT%
pcDNA3.CAT+ cat 22 nts
s as ds
F
Rel. CAT%
pcDNA3.CAT+ cat 23 nts
s as ds
G
Rel. CAT%
pcDNA3.CAT+ egfp 22 nts
s as ds
H
Rel. CAT%
pcDNA3.CAT+ egfp 23 nts
s as ds

Figure 5

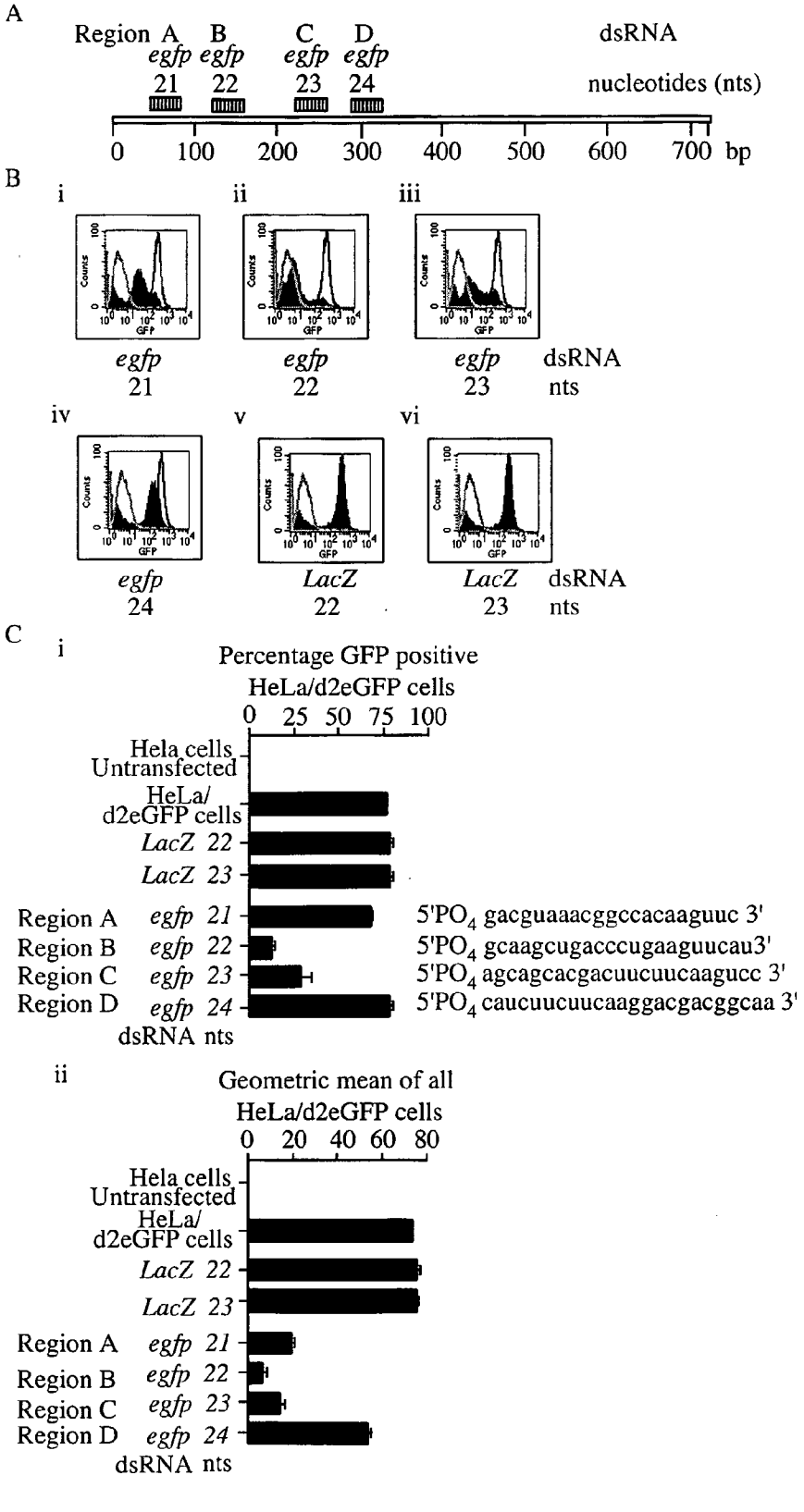
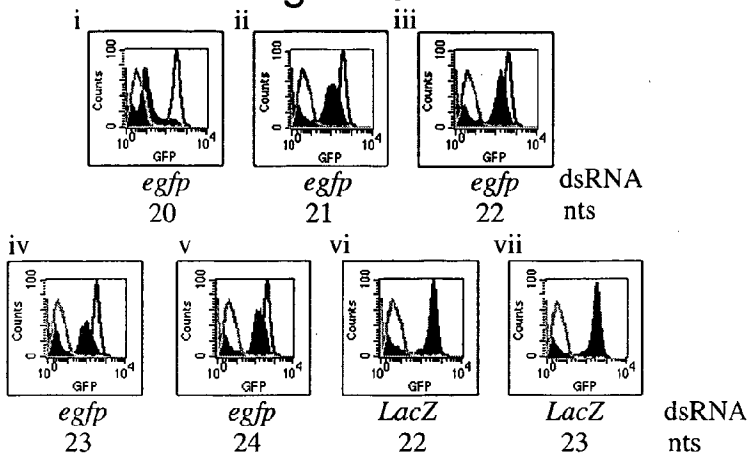
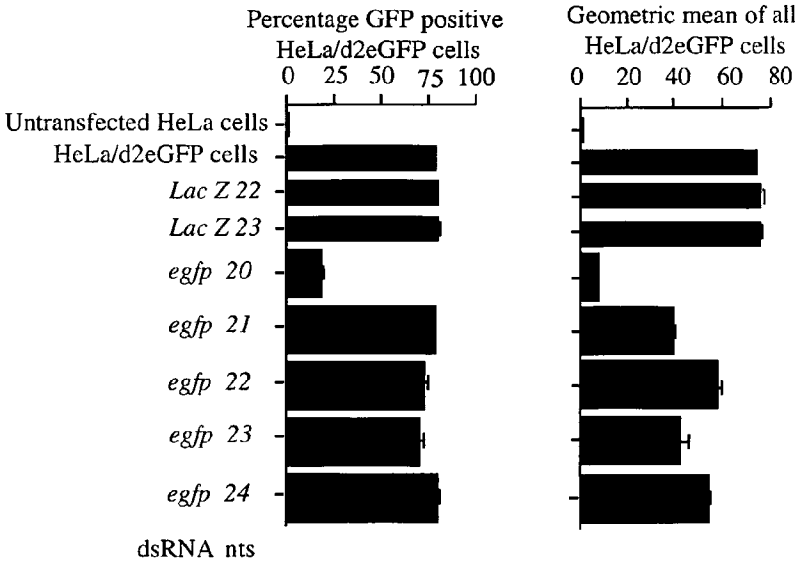

A
Region A B C D
egfp egfp egfp egfp
21 22 23 24
dsRNA
nucleotides (nts)
0 100 200 300 400 500 600 700 bp
B
i ii iii
egfp egfp egfp dsRNA
21 22 23 nts
iv v vi
egfp LacZ LacZ dsRNA
24 22 23 nts
C
i
Percentage GFP positive
HeLa/d2eGFP cells
0 25 50 75 100
Hela cells Untransfected
HeLa/ d2eGFP cells
LacZ 22
LacZ 23
Region A egfp 21 5'PO4 gacguaaacggccacaaguuc 3'
Region B egfp 22 5'PO4 gcaagcugacccugaaguucau3'
Region C egfp 23 5'PO4 agcagcacgacuucuucaagucc 3'
Region D egfp 24 5'PO4 caucuucuucaaggacgacggcaa 3'
dsRNA nts
ii
Geometric mean of all
HeLa/d2eGFP cells
0 20 40 60 80
Hela cells Untransfected
HeLa/ d2eGFP cells
LacZ 22
LacZ 23
Region A egfp 21
Region B egfp 22
Region C egfp 23
Region D egfp 24
dsRNA nts

Figure 6

A



B



egfp 20 5'PO4 r(C A U C U U C U U C A A G G A C G A C G) 3'
3' (u g g u a g a a g a a g u u c c u g c u) r 5'PO4
1 20

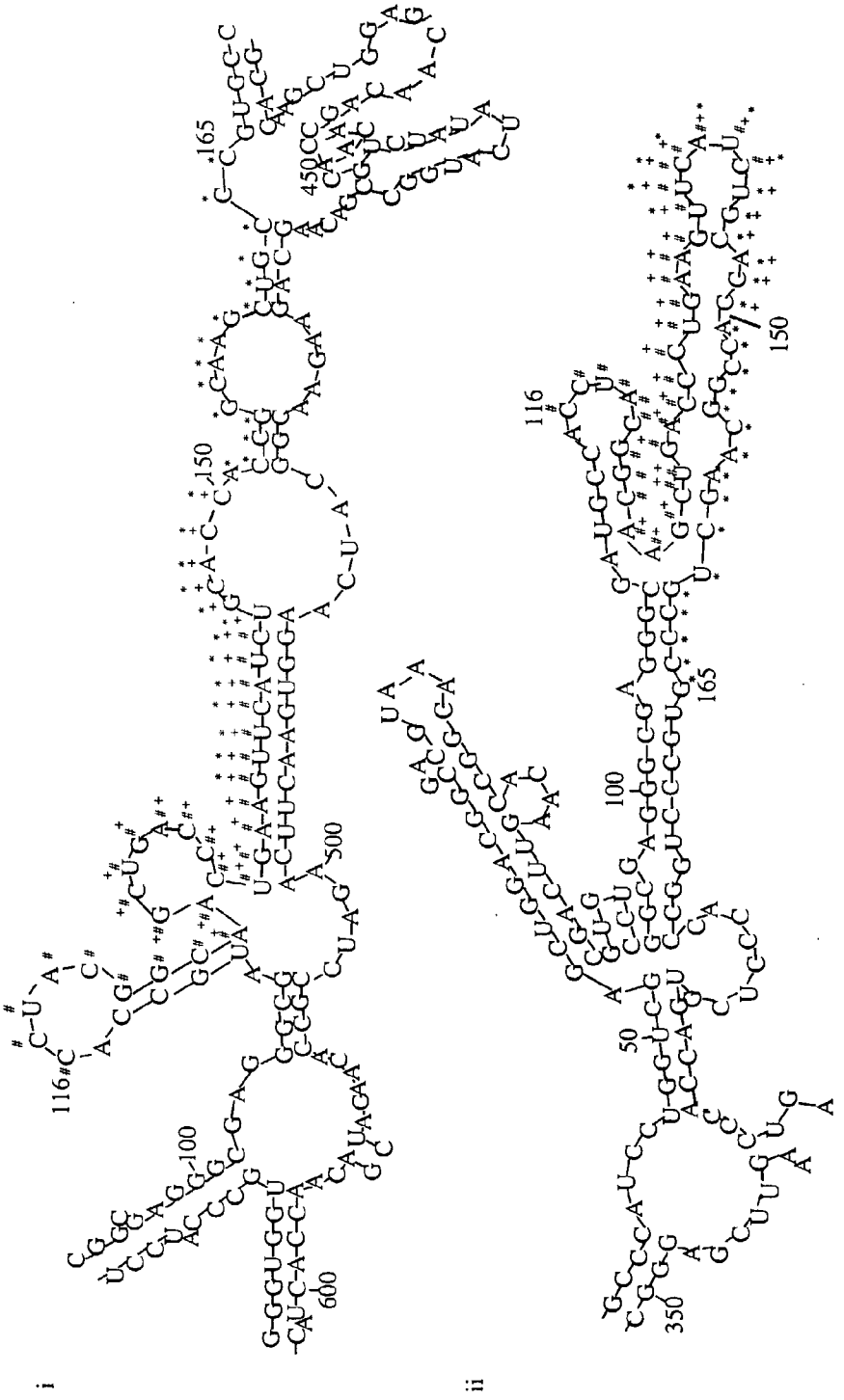
egfp 21 5'PO4 r(C A U C U U C U U C A A G G A C G A C G G) 3'
3' (u g g u a g a a g a a g u u c c u g c u g) r 5'PO4
1 21

egfp 22 5'PO4 r(C A U C U U C U U C A A G G A C G A C G G C) 3'
3' (u g g u a g a a g a a g u u c c u g c u g c) r 5'PO4
1 22

egfp 23 5'PO4 r(C A U C U U C U U C A A G G A C G A C G G C A) 3'
3' (u g g u a g a a g a a g u u c c u g c u g c c) r 5'PO4
1 23

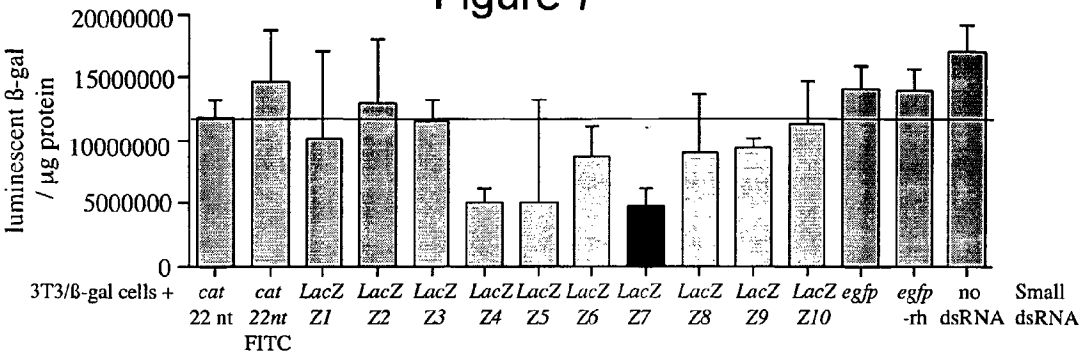
egfp 24 5'PO4 r(C A U C U U C U U C A A G G A C G A C G G C A A) 3'
3' (u g g u a g a a g a a g u u c c u g c u g c c g) r 5'PO4
1 24

Figure 6C



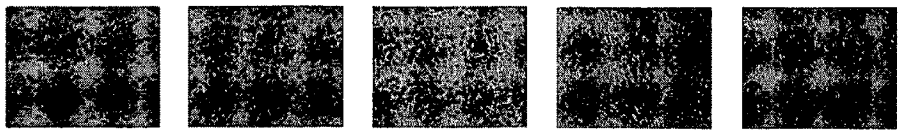
A

Figure 7



B

Repesentative images of 3T3/ß-gal cells 4 days post transfection with the stated siRNA



3T3/ß-gal cells untransfected | 3T3/ß-gal cells + *cat* 22 nt | 3T3/ß-gal cells + *cat*-FITC 22 nt | 3T3/ß-gal cells + *LacZ* Z7 | 3T3/ß-gal cells + *LacZ* Z7 | Small dsRNA

Figure 8

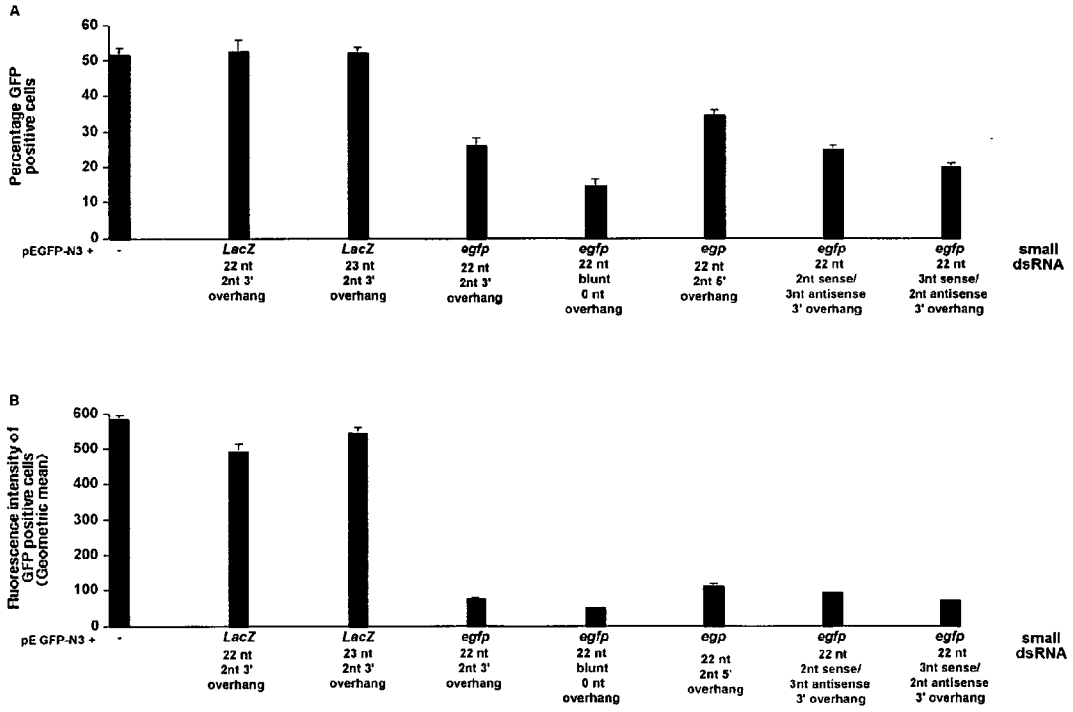


Figure 9

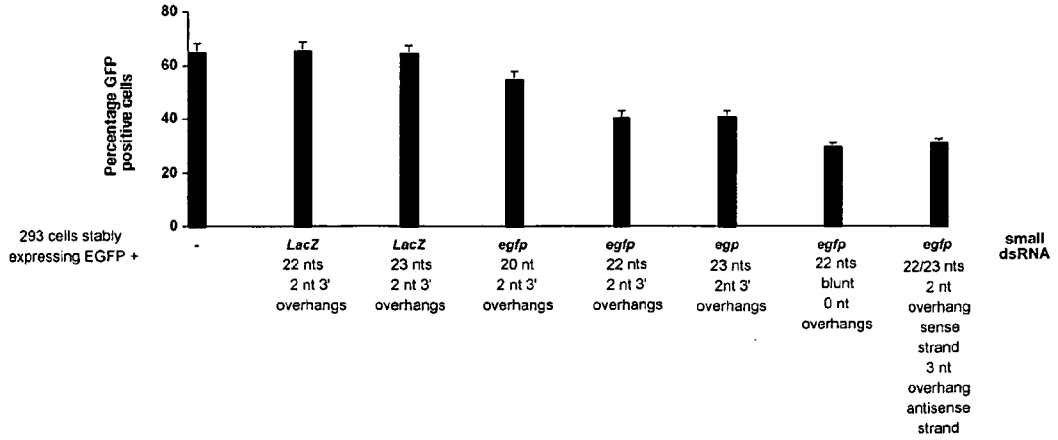


Figure 10

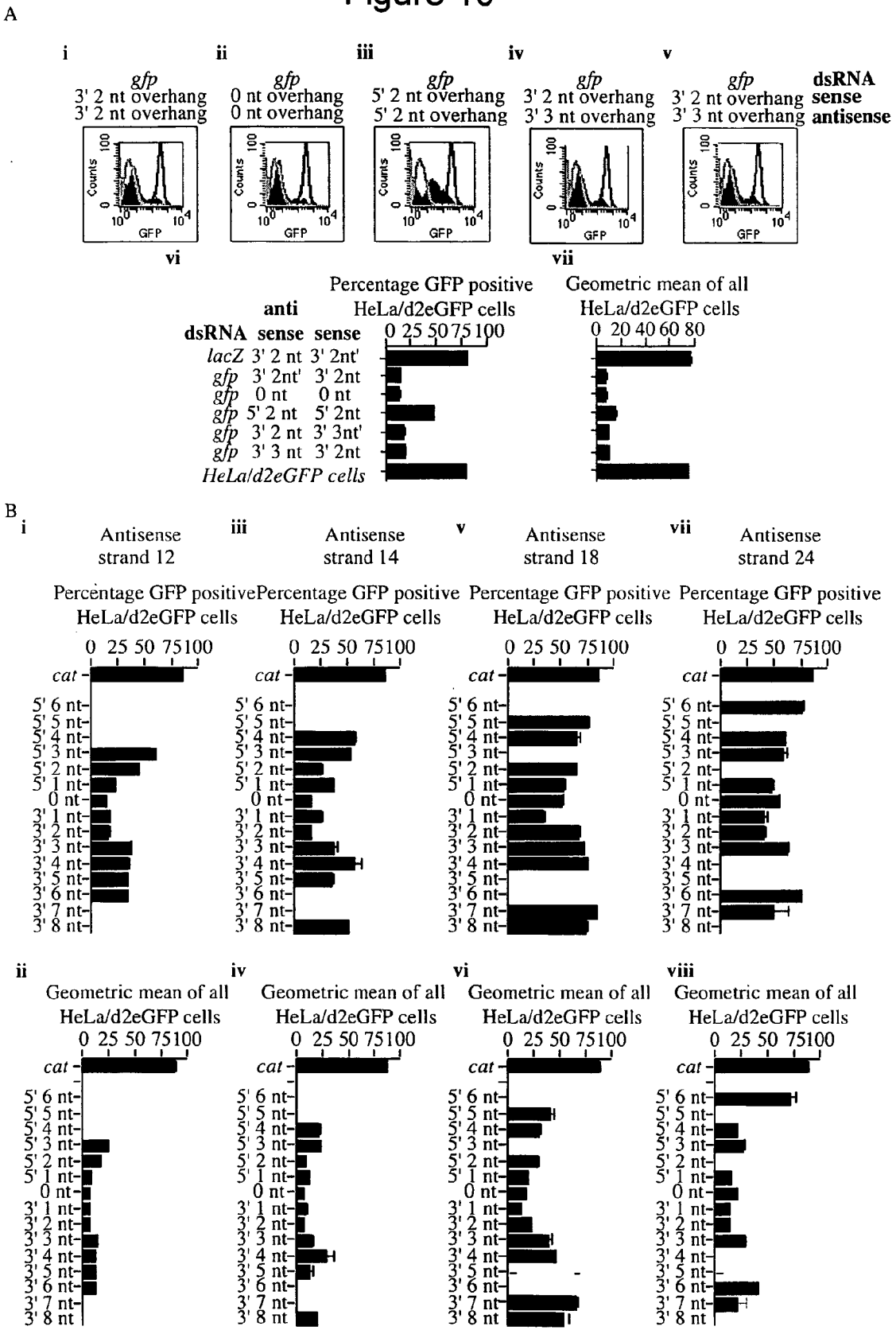

A
i
ii
iii
iv
v
gfp
3' 2 nt overhang
3' 2 nt overhang
gfp
0 nt overhang
0 nt overhang
gfp
5' 2 nt overhang
5' 2 nt overhang
gfp
3' 2 nt overhang
3' 3 nt overhang
gfp
3' 2 nt overhang
3' 3 nt overhang
dsRNA
sense
antisense
Counts
GFP
vi
vii
anti
dsRNA sense sense
lacZ 3' 2 nt 3' 2nt'
gfp 3' 2nt' 3' 2nt
gfp 0 nt 0 nt
gfp 5' 2 nt 5' 2nt
gfp 3' 2 nt 3' 3nt'
gfp 3' 3 nt 3' 2nt
HeLa/d2eGFP cells
Percentage GFP positive HeLa/d2eGFP cells
0 25 50 75 100
Geometric mean of all HeLa/d2eGFP cells
0 20 40 60 80
B
i
Antisense strand 12
iii
Antisense strand 14
v
Antisense strand 18
vii
Antisense strand 24
Percentage GFP positive HeLa/d2eGFP cells
0 25 50 75 100
cat
5' 6 nt
5' 5 nt
5' 4 nt
5' 3 nt
5' 2 nt
5' 1 nt
0 nt
3' 1 nt
3' 2 nt
3' 3 nt
3' 4 nt
3' 5 nt
3' 6 nt
3' 7 nt
3' 8 nt
ii
iv
vi
viii
Geometric mean of all HeLa/d2eGFP cells

Figure 10

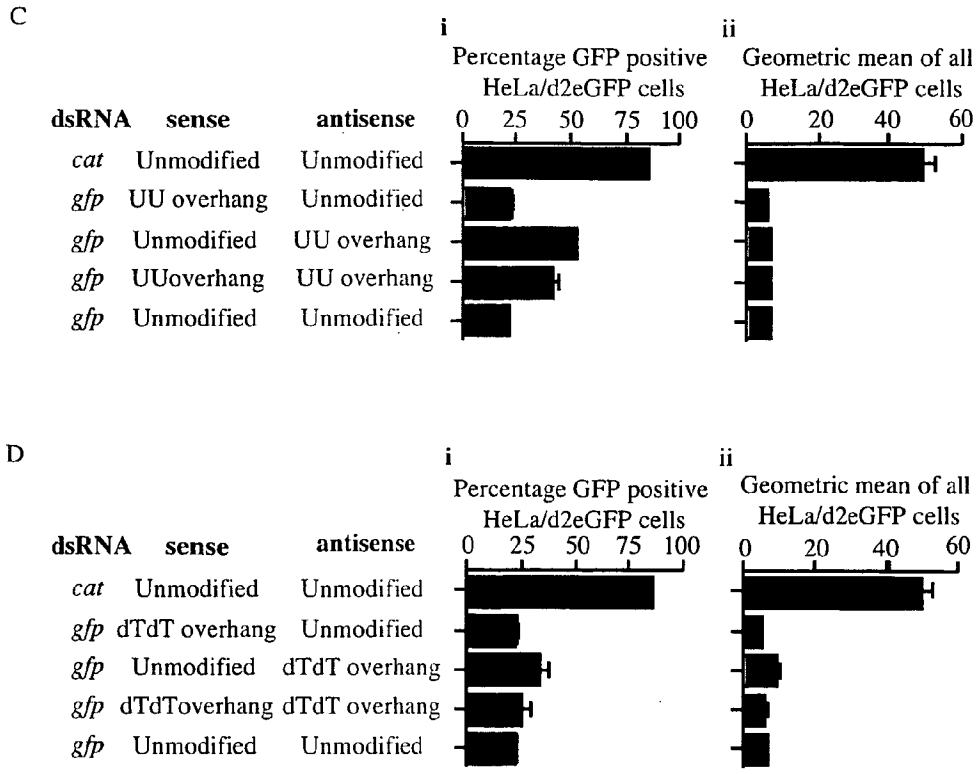
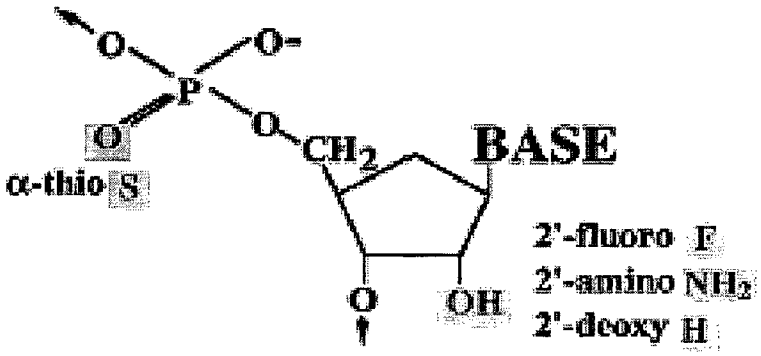

C
i
Percentage GFP positive HeLa/d2eGFP cells
0 25 50 75 100
ii
Geometric mean of all HeLa/d2eGFP cells
0 20 40 60
dsRNA sense antisense
cat Unmodified Unmodified
gfp UU overhang Unmodified
gfp Unmodified UU overhang
gfp UUoverhang UU overhang
gfp Unmodified Unmodified
D
i
Percentage GFP positive HeLa/d2eGFP cells
0 25 50 75 100
ii
Geometric mean of all HeLa/d2eGFP cells
0 20 40 60
dsRNA sense antisense
cat Unmodified Unmodified
gfp dTdT overhang Unmodified
gfp Unmodified dTdT overhang
gfp dTdToverhang dTdT overhang
gfp Unmodified Unmodified

Figure 11

A

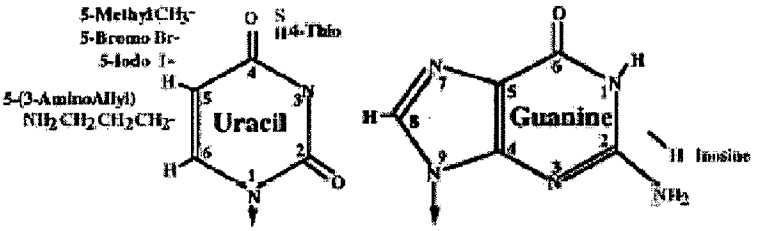


B

Sense Strand of Trigger	Antisense Strand of Trigger	*unc-22* interference
Unmodified RNA	Unmodified RNA	+++
cytidine->α-thiocytidine	cytidine->α-thiocytidine	+++
guanosine->α-thioguanosine	guanosine->α-thioguanosine	+++
adenosine->α-thioadenosine	adenosine->α-thioadenosine	+++
uracil->α-thiouracil	uracil->α-thiouracil	++
uracil->2'-fluorouracil	Unmodified RNA	+++
Unmodified RNA	uracil->2'-fluorouracil	+++
uracil->2'-aminouracil	Unmodified RNA	++
Unmodified RNA	uracil->2'-aminouracil	-
cytidine->2'-aminocytidine	Unmodified RNA	++
Unmodified RNA	cytidine->2'-aminocytidine	-
uracil->2'-deoxythymidine	Unmodified RNA	++
Unmodified RNA	uracil->2'-deoxythymidine	+
cytidine->2'-deoxycytidine	Unmodified RNA	++
Unmodified RNA	cytidine->2'-deoxycytidine	+
Unmodified RNA	DNA	-
DNA	Unmodified RNA	-

Figure 12

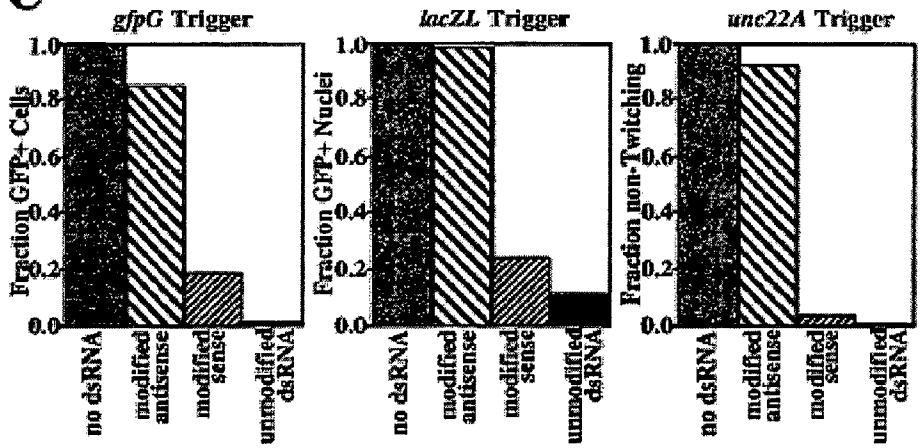
A



B

Sense Strand of Trigger	Antisense Strand of Trigger	*unc-22* Interference
Unmodified RNA	Unmodified RNA	+++
uracil–>4-thiouracil	Unmodified RNA	+++
Unmodified RNA	uracil–>4-thiouracil	+++
uracil–>5-bromouracil	Unmodified RNA	+++
Unmodified RNA	uracil–>5-bromouracil	+++
uracil–>5-iodouracil	Unmodified RNA	++
Unmodified RNA	uracil–>5-iodouracil	+
uracil–>5-(3-aminoallyl)uracil	Unmodified RNA	++
Unmodified RNA	uracil–>5-(3-aminoallyl)uracil	+/-
guanosine–>inosine	Unmodified RNA	+
Unmodified RNA	guanosine–>inosine	+

C



unc54A Trigger

RNA Trigger		progeny as larvae	progeny as adults
no dsRNA	-	motile	motile
modified antisense	30µg/ml	motile	motile
modified sense	60µg/ml	slow	paralyzed
	12µg/ml	slow	slow
	2µg/ml	slow	motile
unmodified dsRNA	50µg/ml	slow	paralyzed
	10µg/ml	slow	slow/paralyzed

Figure 13

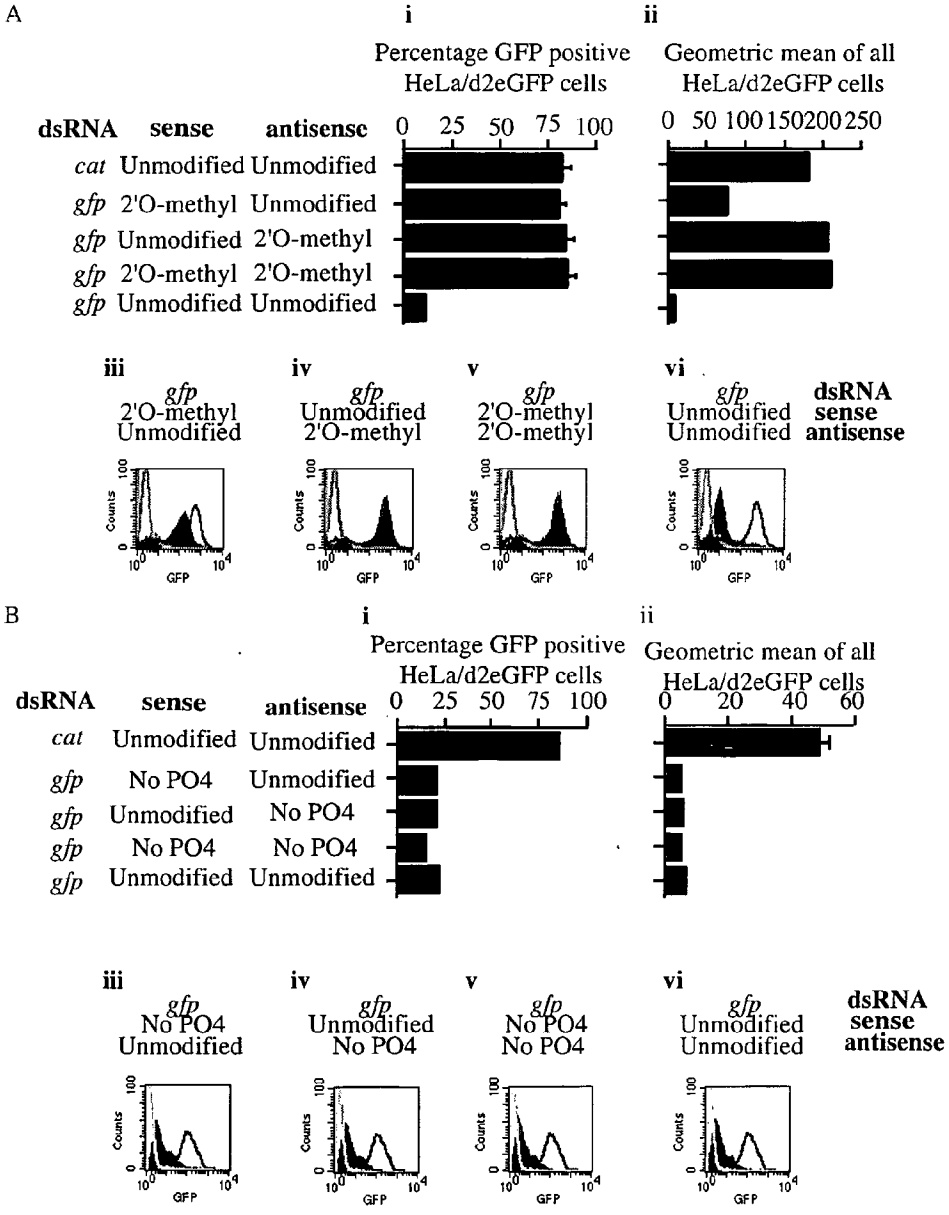

A
i
Percentage GFP positive HeLa/d2eGFP cells
ii
Geometric mean of all HeLa/d2eGFP cells
dsRNA sense antisense
cat Unmodified Unmodified
gfp 2'O-methyl Unmodified
gfp Unmodified 2'O-methyl
gfp 2'O-methyl 2'O-methyl
gfp Unmodified Unmodified
0 25 50 75 100
0 50 100 150 200 250
iii
gfp
2'O-methyl
Unmodified
iv
gfp
Unmodified
2'O-methyl
v
gfp
2'O-methyl
2'O-methyl
vi
gfp
Unmodified
Unmodified
dsRNA
sense
antisense
Counts
GFP
B
i
Percentage GFP positive HeLa/d2eGFP cells
ii
Geometric mean of all HeLa/d2eGFP cells
dsRNA sense antisense
cat Unmodified Unmodified
gfp No PO4 Unmodified
gfp Unmodified No PO4
gfp No PO4 No PO4
gfp Unmodified Unmodified
0 25 50 75 100
0 20 40 60
iii
gfp
No PO4
Unmodified
iv
gfp
Unmodified
No PO4
v
gfp
No PO4
No PO4
vi
gfp
Unmodified
Unmodified
dsRNA
sense
antisense
Counts
GFP

Figure 13

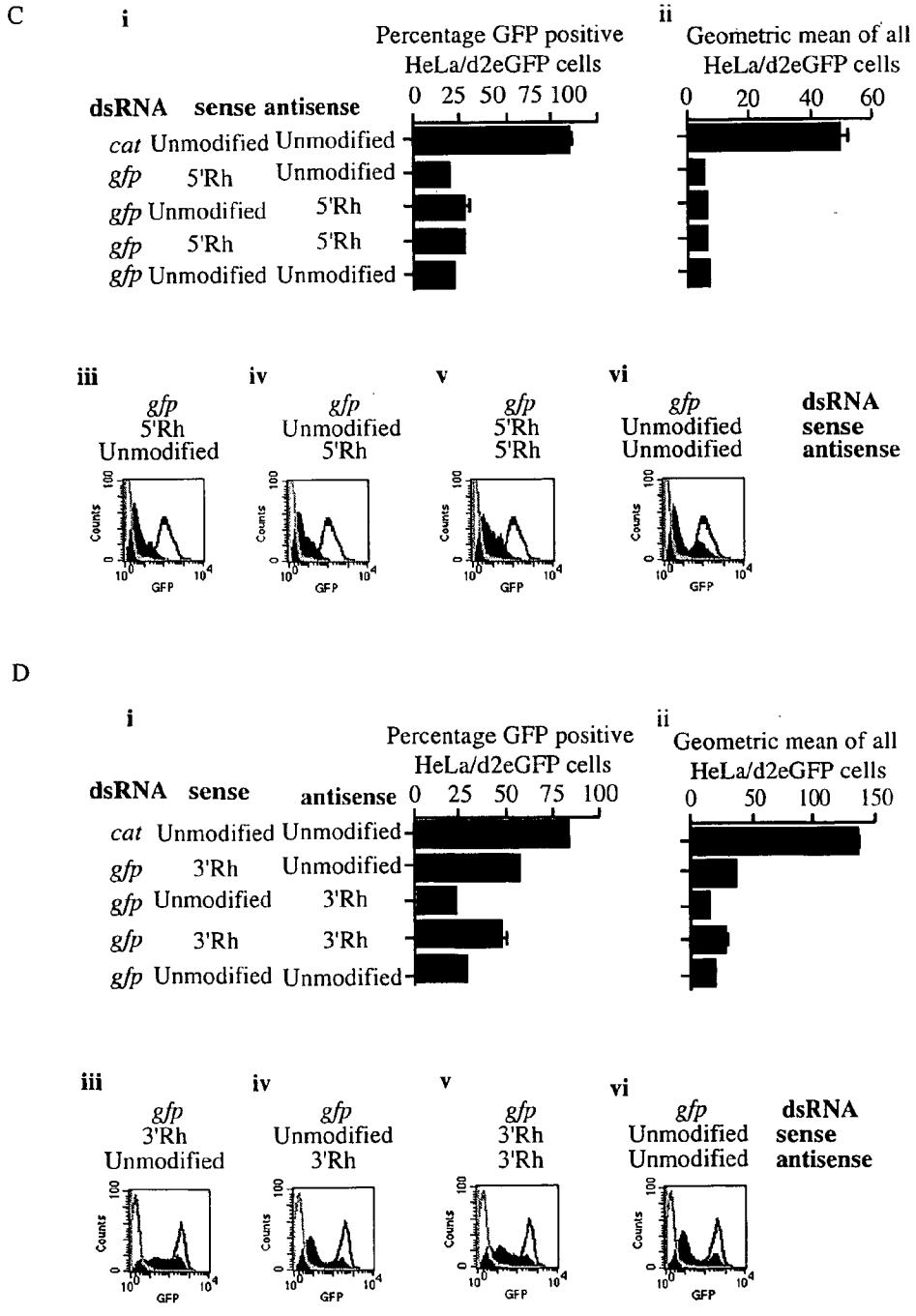

C
i
Percentage GFP positive
HeLa/d2eGFP cells
0 25 50 75 100
dsRNA sense antisense
cat Unmodified Unmodified
gfp 5'Rh Unmodified
gfp Unmodified 5'Rh
gfp 5'Rh 5'Rh
gfp Unmodified Unmodified
ii
Geometric mean of all
HeLa/d2eGFP cells
0 20 40 60
iii
gfp
5'Rh
Unmodified
iv
gfp
Unmodified
5'Rh
v
gfp
5'Rh
5'Rh
vi
gfp
Unmodified
Unmodified
dsRNA
sense
antisense
Counts
GFP
D
i
Percentage GFP positive
HeLa/d2eGFP cells
dsRNA sense antisense 0 25 50 75 100
cat Unmodified Unmodified
gfp 3'Rh Unmodified
gfp Unmodified 3'Rh
gfp 3'Rh 3'Rh
gfp Unmodified Unmodified
ii
Geometric mean of all
HeLa/d2eGFP cells
0 50 100 150
iii
gfp
3'Rh
Unmodified
iv
gfp
Unmodified
3'Rh
v
gfp
3'Rh
3'Rh
vi
gfp
Unmodified
Unmodified
dsRNA
sense
antisense
Counts
GFP

Figure 14

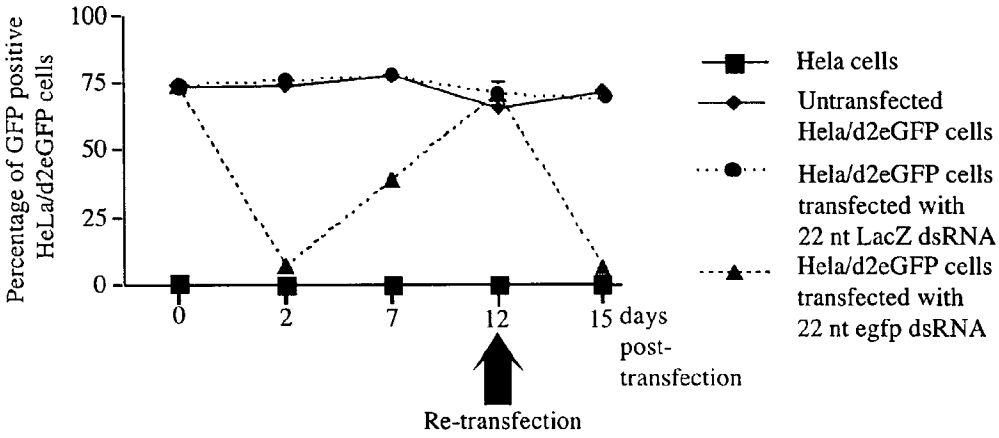

Percentage of GFP positive HeLa/d2eGFP cells
100
75
50
25
0
0
2
7
12
15 days post-transfection
Re-transfection
Hela cells
Untransfected Hela/d2eGFP cells
Hela/d2eGFP cells transfected with 22 nt LacZ dsRNA
Hela/d2eGFP cells transfected with 22 nt egfp dsRNA

Figure 15

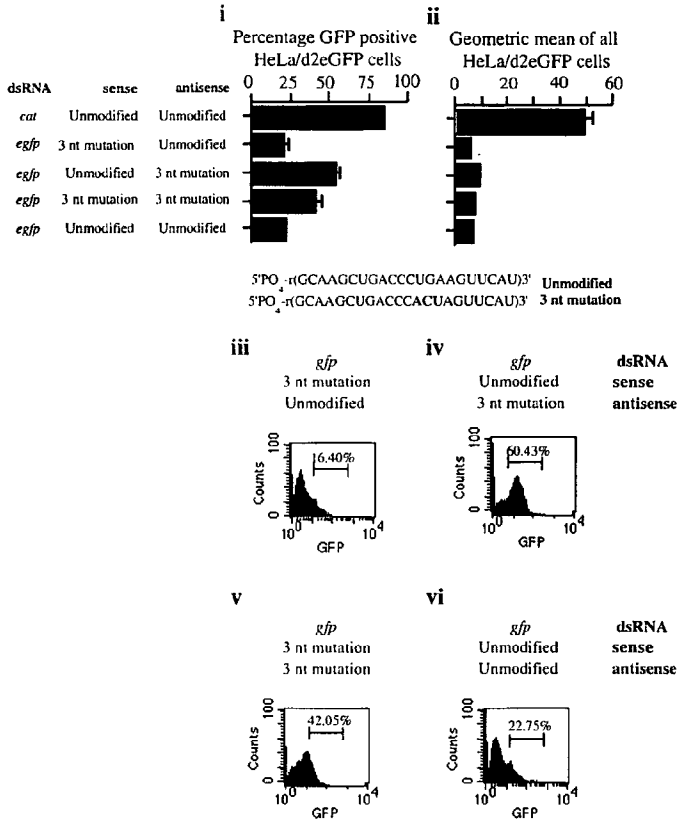

i
Percentage GFP positive HeLa/d2eGFP cells
0 25 50 75 100
ii
Geometric mean of all HeLa/d2eGFP cells
0 20 40 60
dsRNA sense antisense
cat Unmodified Unmodified
egfp 3 nt mutation Unmodified
egfp Unmodified 3 nt mutation
egfp 3 nt mutation 3 nt mutation
egfp Unmodified Unmodified
5'PO4-r(GCAAGCUGACCCUGAAGUUCAU)3' Unmodified
5'PO4-r(GCAAGCUGACCCACUAGUUCAU)3' 3 nt mutation
iii
gfp
3 nt mutation
Unmodified
16.40%
Counts
GFP
iv
gfp
Unmodified
3 nt mutation
dsRNA
sense
antisense
60.43%
Counts
GFP
v
gfp
3 nt mutation
3 nt mutation
42.05%
Counts
GFP
vi
gfp
Unmodified
Unmodified
dsRNA
sense
antisense
22.75%
Counts
GFP

Figure 16

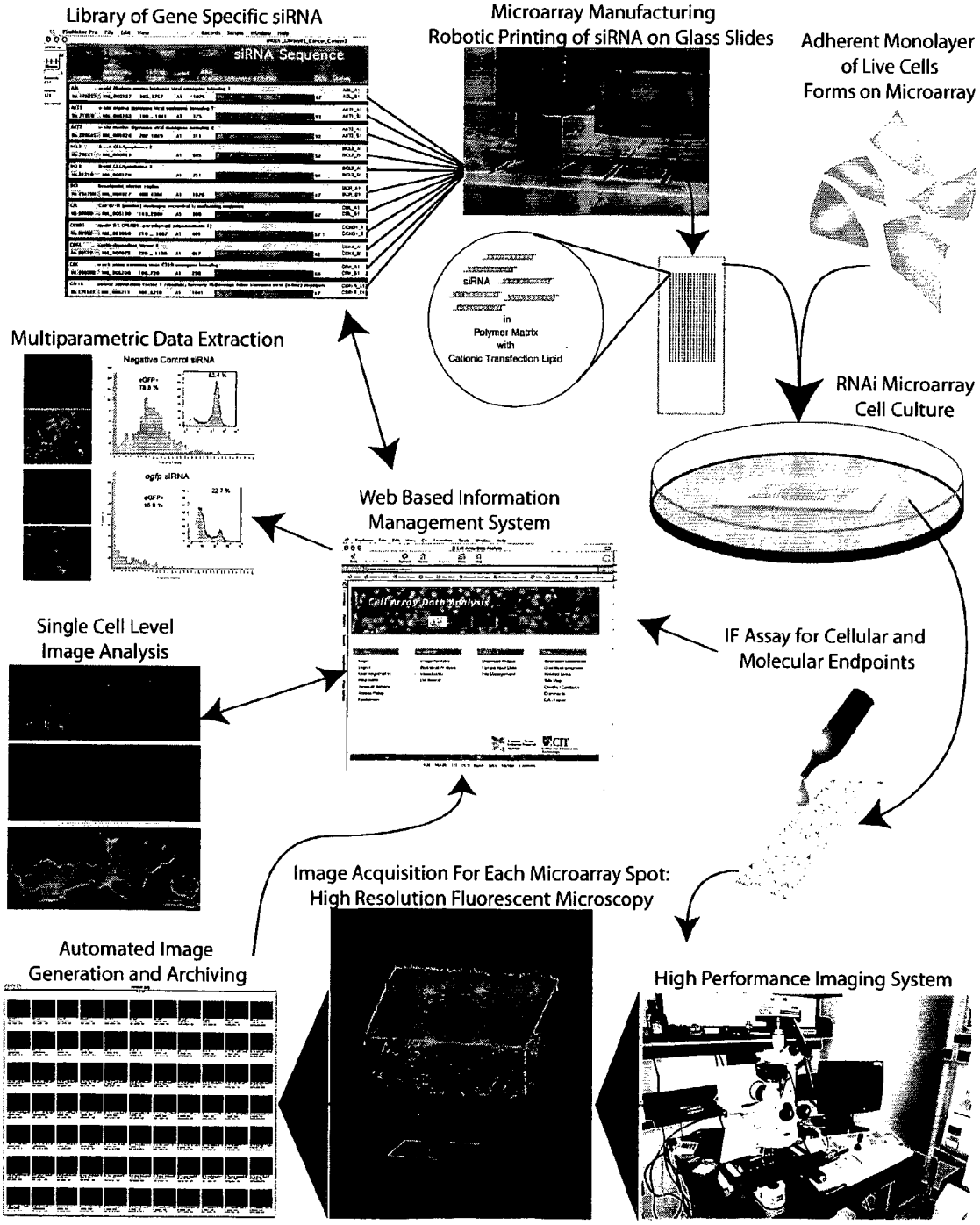

Library of Gene Specific siRNA
siRNA Sequence
Microarray Manufacturing
Robotic Printing of siRNA on Glass Slides
Adherent Monolayer
of Live Cells
Forms on Microarray
siRNA
in
Polymer Matrix
with
Cationic Transfection Lipid
Multiparametric Data Extraction
Negative Control siRNA
ogfp siRNA
RNAi Microarray
Cell Culture
Web Based Information
Management System
Single Cell Level
Image Analysis
IF Assay for Cellular and
Molecular Endpoints
Image Acquisition For Each Microarray Spot:
High Resolution Fluorescent Microscopy
Automated Image
Generation and Archiving
High Performance Imaging System

Figure 17

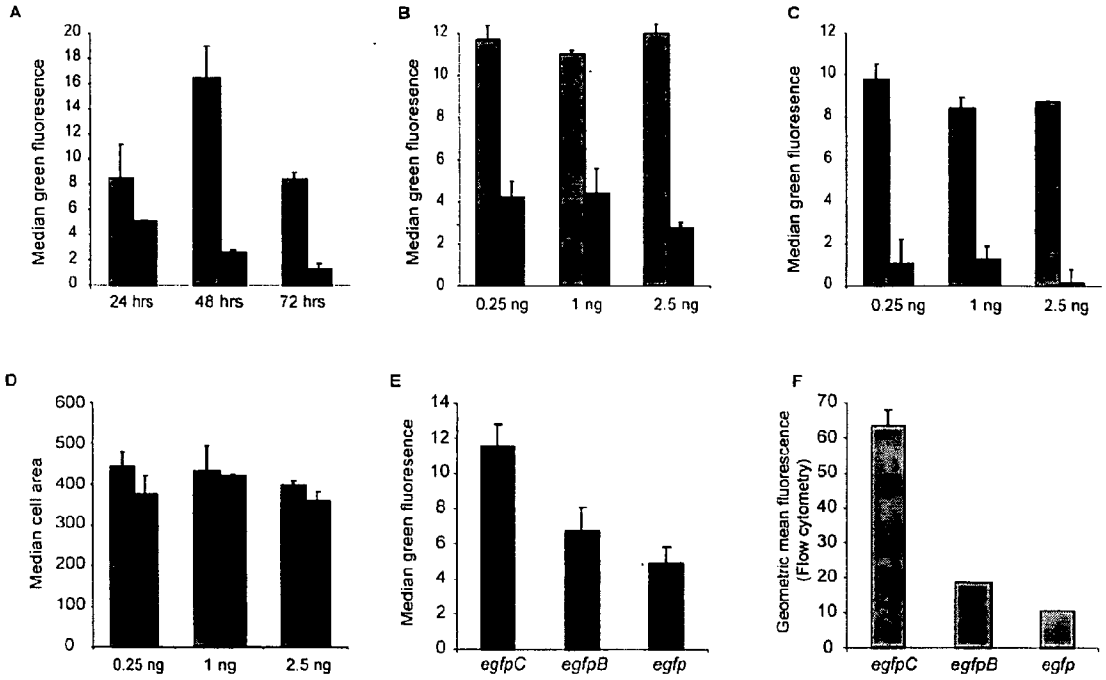

A
Median green fluoresence
20
18
16
14
12
10
8
6
4
2
0
24 hrs
48 hrs
72 hrs
B
Median green fluoresence
12
10
8
6
4
2
0
0.25 ng
1 ng
2.5 ng
C
Median green fluoresence
12
10
8
6
4
2
0
0.25 ng
1 ng
2.5 ng
D
Median cell area
600
500
400
300
200
100
0
0.25 ng
1 ng
2.5 ng
E
Median green fluoresence
14
12
10
8
6
4
2
0
egfpC
egfpB
egfp
F
Geometric mean fluorescence (Flow cytometry)
70
60
50
40
30
20
10
0
egfpC
egfpB
egfp

Figure 18

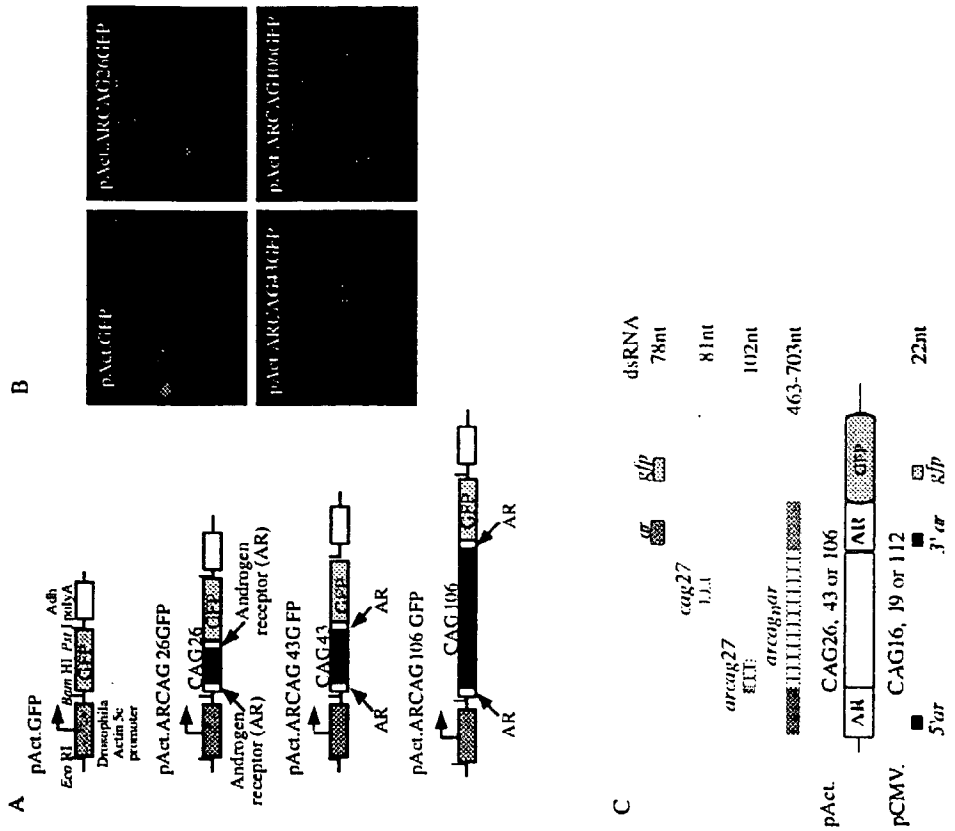

A
pAct.GFP
Eco RI
BamHI
Pst I
Ad. poly A
Drosophila Actin 5c promoter
pAct.ARCAG 26GFP
CAG26
Androgen receptor (AR)
pAct.ARCAG 43GFP
CAG43
AR
pAct.ARCAG 106 GFP
CAG106
B
pAct.GFP
C
dsRNA
78nt
81nt
102nt
463-703nt
22nt
pAct.
pCMV.
CAG26, 43 or 106
CAG16, 19 or 112
AR
GFP
5'ar
3'ar
gfp

Figure 19

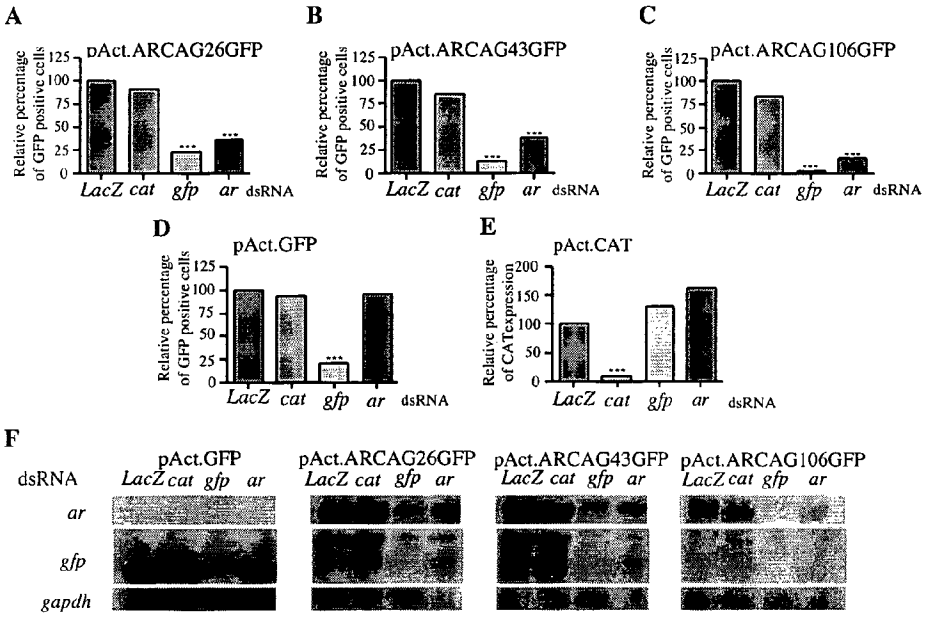

A
pAct.ARCAG26GFP
Relative percentage of GFP positive cells
125
100
75
50
25
0
LacZ cat gfp ar dsRNA
B
pAct.ARCAG43GFP
Relative percentage of GFP positive cells
LacZ cat gfp ar dsRNA
C
pAct.ARCAG106GFP
Relative percentage of GFP positive cells
LacZ cat gfp ar dsRNA
D
pAct.GFP
Relative percentage of GFP positive cells
LacZ cat gfp ar dsRNA
E
pAct.CAT
Relative percentage of CATexpression
200
150
100
50
0
LacZ cat gfp ar dsRNA
F
dsRNA
pAct.GFP
LacZ cat gfp ar
pAct.ARCAG26GFP
LacZ cat gfp ar
pAct.ARCAG43GFP
LacZ cat gfp ar
pAct.ARCAG106GFP
LacZ cat gfp ar
ar
gfp
gapdh

Figure 20

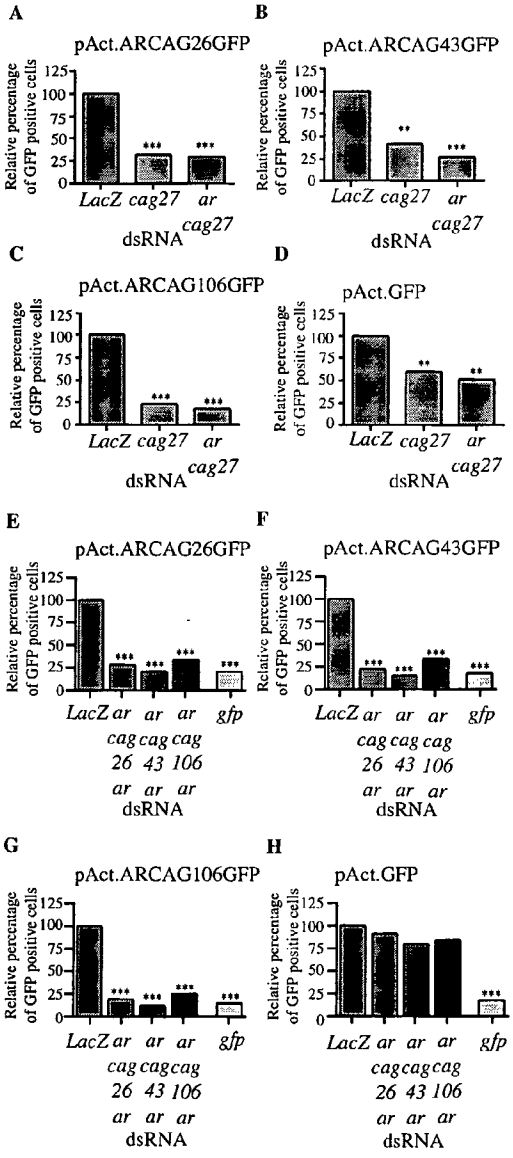

A
pAct.ARCAG26GFP
Relative percentage of GFP positive cells
125
100
75
50
25
0
***
***
LacZ cag27 ar
dsRNA cag27
B
pAct.ARCAG43GFP
Relative percentage of GFP positive cells
**
***
LacZ cag27 ar
dsRNA cag27
C
pAct.ARCAG106GFP
Relative percentage of GFP positive cells
***
***
LacZ cag27 ar
dsRNA cag27
D
pAct.GFP
Relative percentage of GFP positive cells
**
**
LacZ cag27 ar
dsRNA cag27
E
pAct.ARCAG26GFP
Relative percentage of GFP positive cells
***
***
***
***
LacZ ar ar ar gfp
cag cag cag
26 43 106
ar ar ar
dsRNA
F
pAct.ARCAG43GFP
Relative percentage of GFP positive cells
***
***
***
***
LacZ ar ar ar gfp
cag cag cag
26 43 106
ar ar ar
dsRNA
G
pAct.ARCAG106GFP
Relative percentage of GFP positive cells
***
***
***
***
LacZ ar ar ar gfp
cag cag cag
26 43 106
ar ar ar
dsRNA
H
pAct.GFP
Relative percentage of GFP positive cells
***
LacZ ar ar ar gfp
cag cag cag
26 43 106
ar ar ar
dsRNA

Figure 21

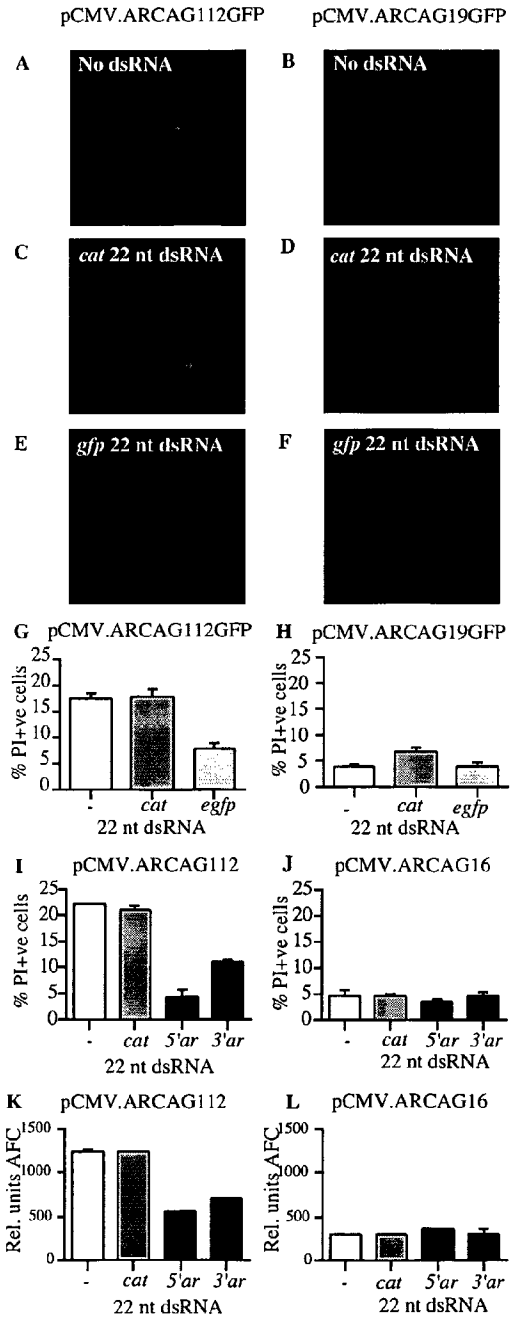

pCMV.ARCAG112GFP
pCMV.ARCAG19GFP
A
No dsRNA
B
No dsRNA
C
cat 22 nt dsRNA
D
cat 22 nt dsRNA
E
gfp 22 nt dsRNA
F
gfp 22 nt dsRNA
G pCMV.ARCAG112GFP
H pCMV.ARCAG19GFP
% PI+ve cells
25
20
15
10
5
0
-
cat
egfp
22 nt dsRNA
I pCMV.ARCAG112
J pCMV.ARCAG16
% PI+ve cells
-
cat
5'ar
3'ar
22 nt dsRNA
K pCMV.ARCAG112
L pCMV.ARCAG16
Rel. units AFC
1500
1000
500
0
-
cat
5'ar
3'ar
22 nt dsRNA

Figure 22A

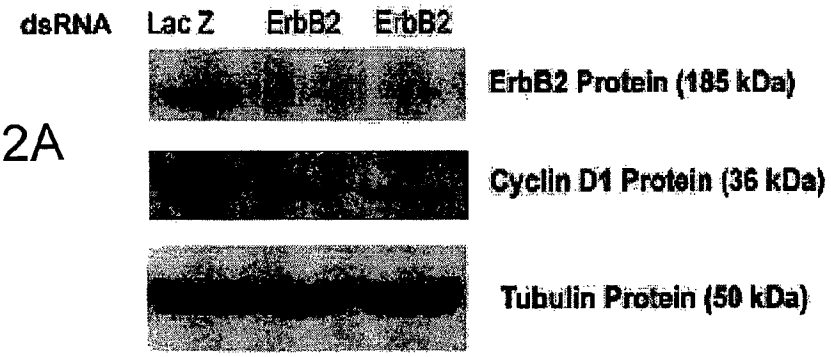


Figure 22B

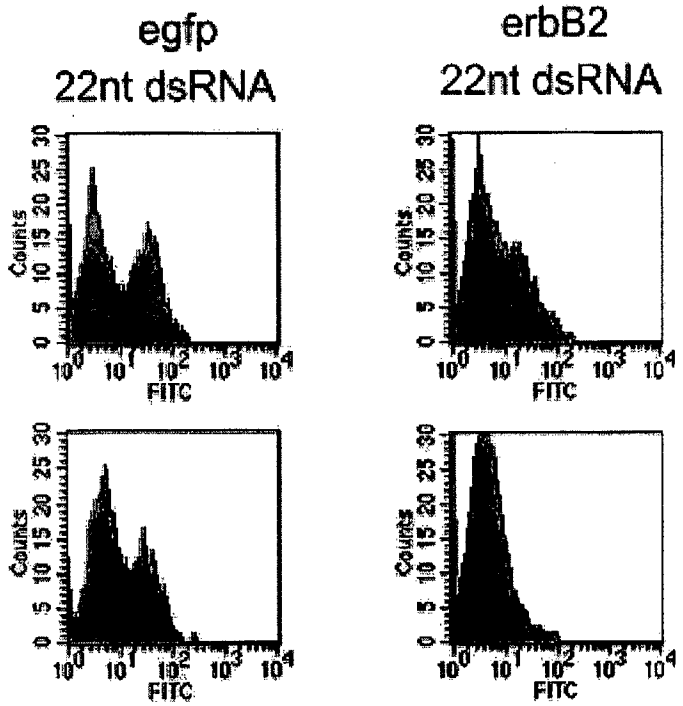


Figure 23

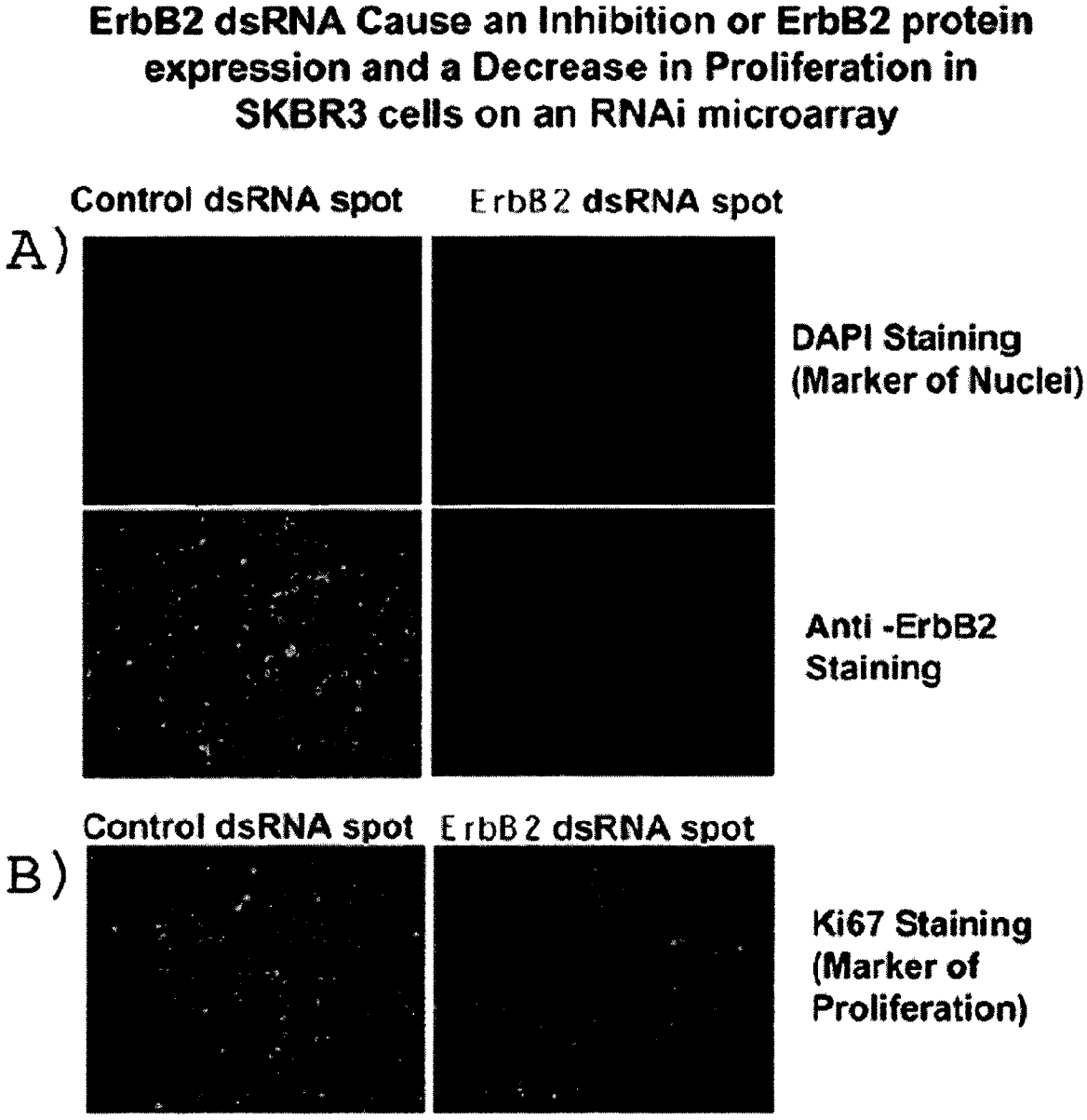

ErbB2 dsRNA Cause an Inhibition or ErbB2 protein expression and a Decrease in Proliferation in SKBR3 cells on an RNAi microarray
Control dsRNA spot
ErbB2 dsRNA spot
A)
DAPI Staining (Marker of Nuclei)
Anti -ErbB2 Staining
Control dsRNA spot
ErbB2 dsRNA spot
B)
Ki67 Staining (Marker of Proliferation)

Figure 24

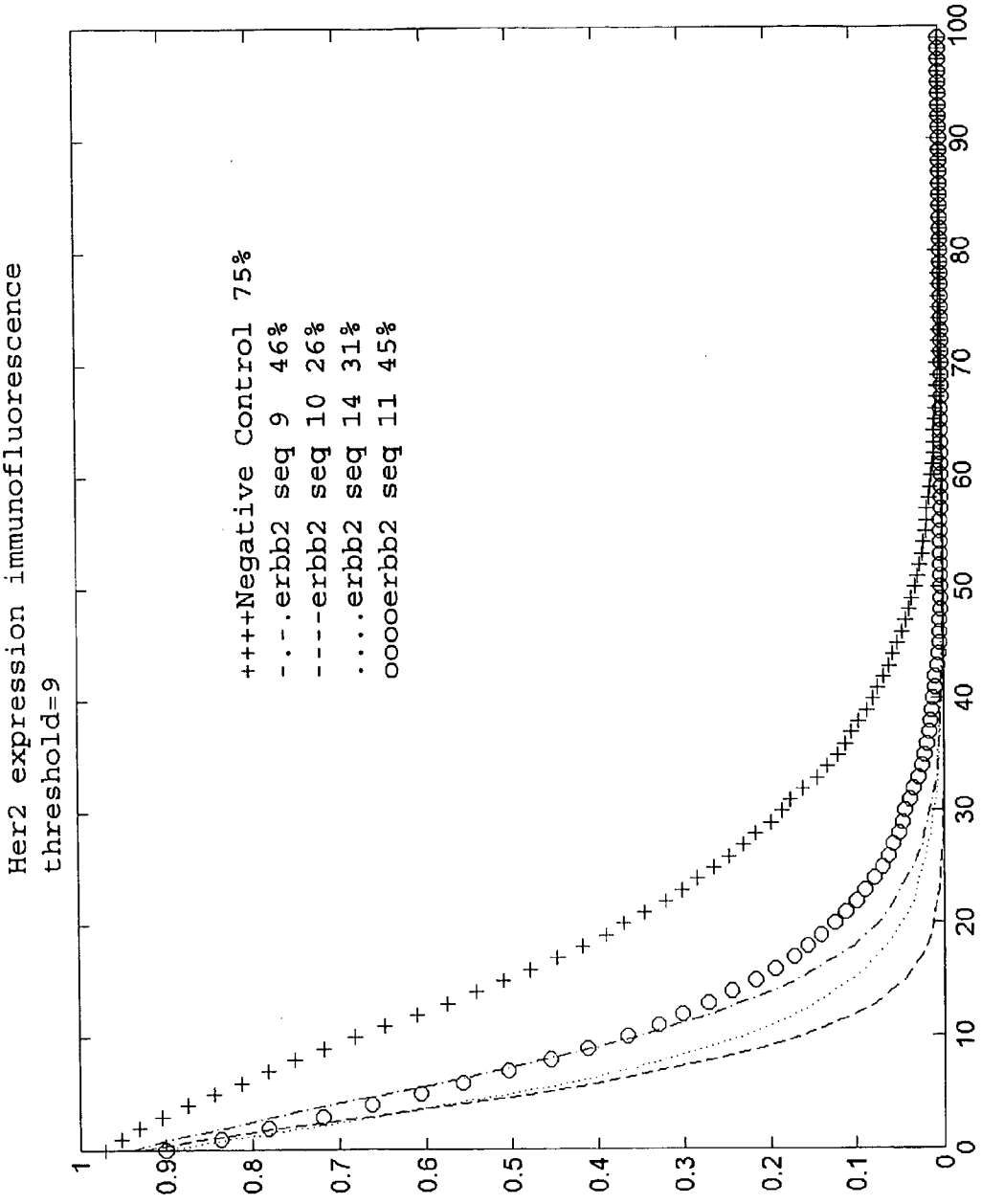

Her2 expression immunofluorescence
threshold=9
++++Negative Control 75%
-.-.erbb2 seq 9 46%
----erbb2 seq 10 26%
....erbb2 seq 14 31%
ooooerbb2 seq 11 45%
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
10
20
30
40
50
60
70
80
90
100

Figure 25

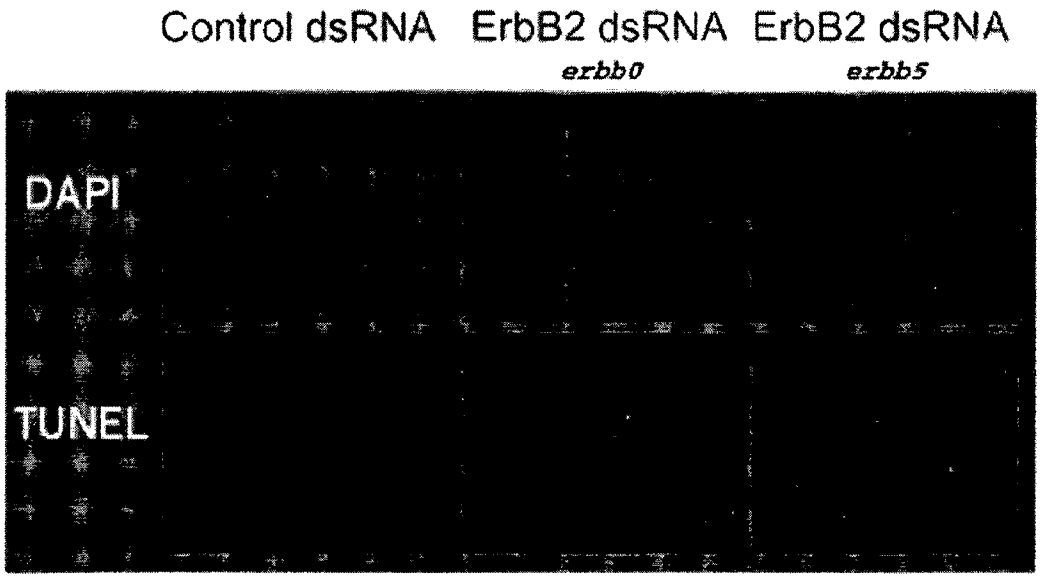

Control dsRNA
ErbB2 dsRNA
erbb0
ErbB2 dsRNA
erbb5
DAPI
TUNEL

Figure 26

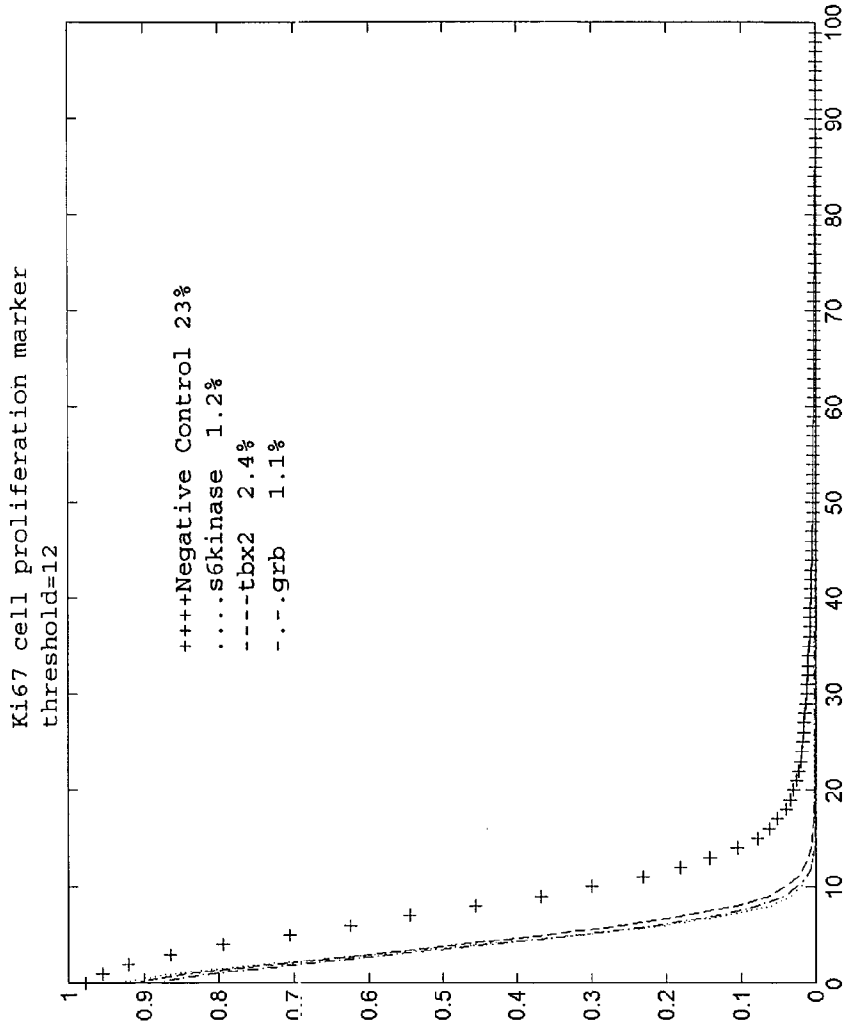
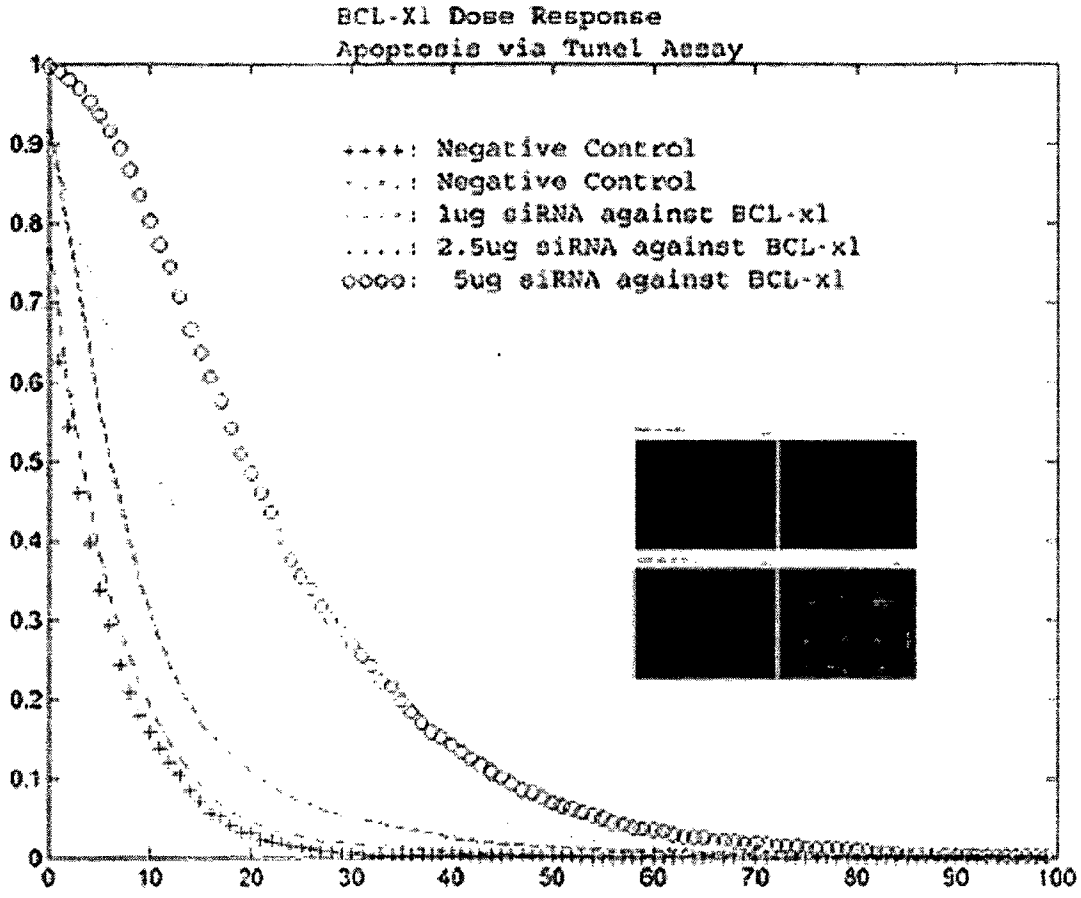

Ki67 cell proliferation marker
threshold=12
++++Negative Control 23%
....s6kinase 1.2%
----tbx2 2.4%
-.-.grb 1.1%
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80
90
100

Figure 27


BCL-Xl Dose Response
Apoptosis via Tunel Assay
++++: Negative Control
....: Negative Control
----: 1ug siRNA against BCL-xl
....: 2.5ug siRNA against BCL-xl
oooo: 5ug siRNA against BCL-xl
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
10
20
30
40
50
60
70
80
90
100

SEQUENCE LISTING

<110> THE GOVERNMENT OF THE UNITED STATES OF AMERICA AS REPRESENTED BY THE SECRETARY OF THE DEPARTMENT OF HEALTH AND HUMAN SERVICES
CARNEGIE INSTITUTE OF WASHINGTON
Caplen, Natasha J.
Morgan, Richard A.
Fire, Andrew
Parrish, Susan
Mousses, Spyro
Kallioniemi, Olli P.
Cornelison, J. Robert

<120> SPECIFIC INHIBITION OF GENE EXPRESSION BY SMALL DOUBLE STRANDED RNAS

<130> 4239-62788

<150> US 60/370,970
<151> 2002-04-08

<150> US 60/308,640
<151> 2001-07-30

<160> 314

<170> PatentIn version 3.1

<210> 1
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 1
gacguaaacg gccacaaguu c 21

<210> 2
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 2
acuuguggcc guuuacgucg c 21

<210> 3
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 3
gcaagcugac ccugaaguuc au 22

<210> 4
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 4
gaacuucagg gucagcuugc cg 22

<210> 5
<211> 23
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 5
agcagcacga cuucuucaag ucc 23

<210> 6
<211> 23
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 6
acuugaagaa gucgugcugc uuc 23

<210> 7
<211> 24
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 7
caucuucuuc aaggacgacg gcaa 24

<210> 8
<211> 24
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 8
gccgucgucc uugaagaaga uggu 24

<210> 9
<211> 25
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 9
gcacaagcug gaguacaacu acaac 25

<210> 10
<211> 25
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 10
uguaguugua cuccagcuuc ugccc 25

<210> 11
<211> 26
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 11
cgacaagcag aagaacggca ucaagg 26

<210> 12
<211> 26
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 12
uugaugccgu ucuucugcuu gucggc 26

<210> 13
<211> 27
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 13
cuucaagauc cgccacaaca ucgagga 27

<210> 14

<211> 27
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 14
cucgauguug uggcggaucu ugaaguu 27

<210> 15
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 15
gagugaauac cacgacgauu uc 22

<210> 16
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 16
aaucgucgug guauucacuc ca 22

<210> 17
<211> 23
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 17
ggagugaaua ccacgacgau uuc 23

<210> 18
<211> 23
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 18
aaucgucgug guauucacuc cag 23

<210> 19
<211> 23

<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 19
ucugucucug cuccucucgg cgg 23

<210> 20
<211> 23
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 20
gccgagagga gcagagacag agg 23

<210> 21
<211> 24
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 21
cucugucucu gcuccucucg gcgg 24

<210> 22
<211> 24
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 22
gccgagagga gcagagacag agga 24

<210> 23
<211> 25
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 23
ccucugucuc ugcuccucuc ggcgg 25

<210> 24
<211> 25
<212> RNA

<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 24
gccgagagga gcagagacag aggau 25

<210> 25
<211> 26
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 25
uccucugucu cugcuccucu cggcgg 26

<210> 26
<211> 27
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 26
aacccaucuc caccaaauaa cuacaau 27

<210> 27
<211> 32
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 27
aaaaagcaag ccaacgaaac caaagggacc aa 32

<210> 28
<211> 37
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 28
aaguugaagg uuaagaauga guugggagag gaugaag 37

<210> 29
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 29
gguggcgcug gaugguaagc cg 22

<210> 30
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 30
gcuuaccauc cagcgccacc au 22

<210> 31
<211> 23
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 31
gguggcgcug gaugguaagc cgc 23

<210> 32
<211> 23
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 32
ggcuuaccau ccagcgccac cau 23

<210> 33
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 33
aguugaacuu cagggucagc uu 22

<210> 34
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 34
augaacuuca gggucagcuu gc 22

<210> 35
<211> 23
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 35
gaacuucagg gucagcuugc cgu 23

<210> 36
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 36
aacuucaggg ucagcuugcc g 21

<210> 37
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 37
caucuucuuc aaggacgacg 20

<210> 38
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 38
ucguccuuga agaagauggu 20

<210> 39
<211> 78
<212> RNA
<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 39
cuacguccag gagcgcacca ucuucuucaa ggacgacggc aacuacaaga cccgcgccga 60

ggugaaguuc gagggcga 78

<210> 40
<211> 78
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 40
ucgcccucga acuucaccuc ggcgcggguc uuguaguugc cgucguccuu gaagaagaug 60

gugcgcuccu ggacguag 78

<210> 41
<211> 78
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 41
caucgcucug gagugaauac cacgacgauu uccggcaguu ucuacacaua uauucgcaag 60

auguggcgug uuacggug 78

<210> 42
<211> 78
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 42
caccguaaca cgccacaucu ugcgaauaua uguguagaaa cugccggaaa ucgucguggu 60

auucacucca gagcgaug 78

<210> 43
<211> 81
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 43
gugcucggaa aaccaucuag cccauuggga ccuuuggaag ugucgaaugu cuacgaagau 60

cgcgcagauu uggaguggaa a 81

<210> 44
<211> 85
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 44
guacuuucca cuccaaaucu gcgcgaucuu cguagacauu cgacacuucc aaaggcccca 60

augggcuaga ugguuuuccg agcac 85

<210> 45
<211> 81
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 45
agcuccugca cuggauggug gcgcuggaug guaagccgcu ggcaagcggu gaagugccuc 60

uggaugucgc uccacaaggu a 81

<210> 46
<211> 81
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 46
uaccuugugg agcgacaucc agaggcacuu caccgcuugc cagcggcuua ccauccagcg 60

ccaccaucca gugcaggagc u 81

<210> 47
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 47
aggcacccag cucuuugagg ac 22

<210> 48
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 48
ccucaaagag cugggugccu cg 22

<210> 49
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 49
aacuaugccc uggccgugcu ag 22

<210> 50
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 50
agcacggcca gggcauaguu gu 22

<210> 51
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 51
acaauggaga cccgcugaac aa 22

<210> 52
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 52
guucagcggg ucuccauugu cu 22

<210> 53
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 53
uaccaccccu gucacagggg cc 22

<210> 54
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 54
ccccugugac agggguggua uu 22

<210> 55
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 55
uccccaggag gccugcggga gc 22

<210> 56
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 56
ucccgcaggc cuccugggga gg 22

<210> 57
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 57
ugcagcuucg aagccucaca ga 22

<210> 58
<211> 22
<212> RNA
<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 58
ugugaggcuu cgaagcugca gc 22

<210> 59
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 59
gaucuugaaa ggaggggucu ug 22

<210> 60
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 60
agaccccucc uuucaagauc uc 22

<210> 61
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 61
auccagcgga acccccagcu cu 22

<210> 62
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 62
agcugggggu uccgcuggau ca 22

<210> 63
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 63
gcuaccagga cacgauuuug ug 22

<210> 64
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 64
caaaaucgug uccugguagc ag 22

<210> 65
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 65
gaacaauacc accccuguca ca 22

<210> 66
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 66
ugacaggggu gguauuguuc ag 22

<210> 67
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 67
uggggagaga guucugagga uu 22

<210> 68
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 68
uccucagaac ucucucccca gc 22

<210> 69
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 69
ccagugccaa uauccaggag uu 22

<210> 70
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 70
cuccuggaua uuggcacugg ua 22

<210> 71
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 71
aguaauccgg ggacgaauuc ug 22

<210> 72
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 72
gaauucgucc ccggauuacu ug 22

<210> 73
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 73

uccaguggcc aucaaagugu ug 22

<210> 74
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 74
acacuuugau ggccacugga au 22

<210> 75
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 75
agacgaagca uacgugaugg cu 22

<210> 76
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 76
ccaucacgua ugcuucgucu aa 22

<210> 77
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 77
acugugguggg agcugaugac uu 22

<210> 78
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 78
gucaucagcu cccacacagu ca 22

<210> 79
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 79
cuggcuccga uguauuugau gg 22

<210> 80
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 80
aucaaauaca ucggagccag cc 22

<210> 81
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 81
uggggucguc aaagacguuu 20

<210> 82
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 82
acgucuuuga cgaccccauu 20

<210> 83
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 83
ccuguccuaa ggaaccuucc uu 22

<210> 84
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 84
ggaagguucc uuaggacagg uu 22

<210> 85
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 85
uggugucagu auccaggcuu 20

<210> 86
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 86
gccuggauac ugacaccauu 20

<210> 87
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 87
gugugcaccg gcacagacau g 21

<210> 88
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 88
ugucugugcc ggugcacacu u 21

<210> 89
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 89
guuaccagug ccaauaucca g 21

<210> 90
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 90
ggauauuggc acugguaacu g 21

<210> 91
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 91
aguacacgau gcggagacug c 21

<210> 92
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 92
agucuccgca ucguguacuu c 21

<210> 93
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 93
auugacucug aaugucggcc a 21

<210> 94

<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 94
gccgacauuc agagucaauc a 21

<210> 95
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 95
uacagcggua cagugaggac c 21

<210> 96
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 96
uccucacugu accgcuguag a 21

<210> 97
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 97
cccugaaguu caucugcacc ac 22

<210> 98
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 98
ggugcagaug aacuucaggg uc 22

<210> 99
<211> 22

<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 99
caucugcacc accggcaagc ug 22

<210> 100
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 100
gcuugccggu ggugcagaug aa 22

<210> 101
<211> 23
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 101
caucuucuuc aaggacgacg gca 23

<210> 102
<211> 23
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 102
ccgucguccu ugaagaagau ggu 23

<210> 103
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 103
caucuucuuc aaggacgacg gc 22

<210> 104
<211> 22
<212> RNA

<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 104
cgucguccuu gaagaagaug gu 22

<210> 105
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 105
caucuucuuc aaggacgacg g 21

<210> 106
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 106
gucguccuug aagaagaugg u 21

<210> 107
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 107
cggcaagcug acccugaagu uc 22

<210> 108
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 108
acuucagggu cagcuugccg ua 22

<210> 109
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 109
aucugcacca ccggcaagcu gc 22

<210> 110
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 110
agcuugccgg uggugcagau ga 22

<210> 111
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 111
caagcugacc cugaaguuca uc 22

<210> 112
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 112
ugaacuucag ggucagcuug cc 22

<210> 113
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 113
aagcugaccc ugaaguucau cu 22

<210> 114
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 114
augaacuuca gggucagcuu gc 22

<210> 115
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 115
ucaucugcac caccggcaag cu 22

<210> 116
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 116
cuugccggug gugcagauga ac 22

<210> 117
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 117
agcugacccu gaaguucauc ug 22

<210> 118
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 118
gaugaacuuc agggucagcu ug 22

<210> 119
<211> 22
<212> RNA
<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 119
uacggcaagc ugacccugaa gu 22

<210> 120
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 120
uucaggguca gcuugccgua gg 22

<210> 121
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 121
acggcaagcu gacccugaag uu 22

<210> 122
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 122
cuucaggguc agcuugccgu ag 22

<210> 123
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 123
gcugacccug aaguucaucu gc 22

<210> 124
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 124
agaugaacuu cagggucagc uu 22

<210> 125
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 125
cugacccuga aguucaucug ca 22

<210> 126
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 126
cagaugaacu ucagggucag cu 22

<210> 127
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 127
ggcaagcuga cccugaaguu ca 22

<210> 128
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 128
aacuucaggg ucagcuugcc gu 22

<210> 129
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 129
cugcaccacc ggcaagcugc cc 22

<210> 130
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 130
gcagcuugcc gguggugcag au 22

<210> 131
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 131
acccugaagu ucaucugcac ca 22

<210> 132
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 132
gugcagauga acuucagggu ca 22

<210> 133
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 133
gcaagcugac ccugaaguuc uu 22

<210> 134
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 134

gaacuucagg gucagcuugc uu 22

<210> 135
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (21)..(22)
<223> n = deoxythymine

<400> 135
gcaagcugac ccugaaguuc nn 22

<210> 136
<211> 23
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (22)..(23)
<223> n = deoxythymine

<400> 136
gaacuucagg gucagcuugc unn 23

<210> 137
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (1)..(1)
<223> 2'-hydroxyl replaced by 2'O-methyl

<400> 137
gcaagcugac ccugaaguuc au 22

<210> 138
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (1)..(1)
<223> 2'-hydroxyl replaced by 2'O-methyl

<400> 138
gaacuucagg gucagcuugc cg 22

<210> 139
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (1)..(1)
<223> 5'PO4 omitted

<400> 139
gcaagcugac ccugaaguuc au 22

<210> 140
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (1)..(1)
<223> 5'PO4 omitted

<400> 140
gaacuucagg gucagcuugc cg 22

<210> 141
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (1)..(1)

<223> Rhodamine tagged

<400> 141
gcaagcugac ccugaaguuc au 22

<210> 142
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (1)..(1)
<223> Rhodamine tagged

<400> 142
gaacuucagg gucagcuugc cg 22

<210> 143
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (22)..(22)
<223> Rhodamine tagged

<400> 143
gcaagcugac ccugaaguuc au 22

<210> 144
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (22)..(22)
<223> Rhodamine tagged

<400> 144
gaacuucagg gucagcuugc cg 22

<210> 145
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 145
gcaagcugac ccacuaguuc au 22

<210> 146
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 146
gaacuagugg gucagcuugc cg 22

<210> 147
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (22)..(22)
<223> FITC tagged

<400> 147
gagugaauac cacgacgauu uc 22

<210> 148
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 148
acccuggcgu uacccaacuu aa 22

<210> 149
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 149
aaguuugggua acgccagggu uu 22

<210> 150
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 150
gcuggcugga gugcgaucuu 20

<210> 151
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 151
gaucgcacuc cagccagcuu 20

<210> 152
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 152
ccuaucccau uacggucaau cc 22

<210> 153
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 153
auugaccgua augggauagg 20

<210> 154
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 154
ccgacuacac aaaucagcga uu 22

<210> 155
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 155
ucgcugauuu guguagucgg uu 22

<210> 156
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 156
guucagaugu gcggcgaguu 20

<210> 157
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 157
cucgccgcac aucugaacuu 20

<210> 158
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 158
cuuuaacgcc gugcgcuguu 20

<210> 159
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 159

cagcgcacgg cguuaaaguu 20

<210> 160
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 160
gccaauauug aaacccacgg 20

<210> 161
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 161
gugggguuuca auauuggcuu 20

<210> 162
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 162
cugugccgaa augguccauc aa 22

<210> 163
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 163
gauggaccau uucggcacag cc 22

<210> 164
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 164
gcaaaacacc agcagcaguu 20

<210> 165
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 165
cugcugcugg uguuuugcuu 20

<210> 166
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 166
gugaccagcg aauaccuguu 20

<210> 167
<211> 20
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 167
cagguauucg cuggucacuu 20

<210> 168
<211> 78
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 168
agggugagga ugguucuccc caagcccauc guagaggccc cacaggcuac cugguccugg 60

augaggaaca gcaaccuu 78

<210> 169
<211> 78
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 169

aagguugcug uuccucaucc aggaccaggu agccuguggg gccucuacga ugggcuuggg 60

gagaaccauc cucacccu 78

<210> 170
<211> 81
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 170
cagcagcagc agcagcagca gcagcagcag cagcagcagc agcagcagca gcagcagcag 60

cagcagcagc agcagcagca g 81

<210> 171
<211> 81
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 171
cugcugcugc ugcugcugcu gcugcugcug cugcugcugc ugcugcugcu gcugcugcug 60

cugcugcugc ugcugcugcu g 81

<210> 172
<211> 104
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 172
ggcgccaguu ugcugcugcu gcagcagcag cagcagcagc agcagcagca gcagcagcag 60

cagcagcagc agcagcagca gcagcagcag cagcagcagc agca 104

<210> 173
<211> 104
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 173
ugcugcugcu gcugcugcug cugcugcugc ugcugcugcu gcugcugcug cugcugcugc 60

ugcugcugcu gcugcugcug cugcagcagc agcaaacugg cgcc 104

<210> 174
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 174
gcucaaggau ggaagugcag uu 22

<210> 175
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 175
cugcacuucc auccuugagc uu 22

<210> 176
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 176
cuggaugagg aacagcaacc uu 22

<210> 177
<211> 22
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 177
gguugcuguu ccucauccag ga 22

<210> 178
<211> 23
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (22)..(23)
<223> n = deoxythymine

<400> 178
aagugugcac cggcacagac ann 23

<210> 179
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (20)..(21)
<223> n = deoxythymine

<400> 179
ugucugugcc ggugcacacn n 21

<210> 180
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 180
guggacaacu acauccgcuc c 21

<210> 181
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 181
agcggaugua guuguccacg a 21

<210> 182
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 182
gagguugcuu caugugcgaa a 21

<210> 183
<211> 21

<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 183
ucgcacauga agcaaccuca c 21

<210> 184
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 184
cuacaaggac uacauccggg a 21

<210> 185
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 185
ccggauguag uccuuguaga u 21

<210> 186
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 186
uucgaugaca gcuccaacgu c 21

<210> 187
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 187
cguuggagcu gucaucgaag u 21

<210> 188
<211> 21
<212> RNA

<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (20)..(21)
<223> n = deoxythymine

<400> 188
ugaacucuuc cgggaugggn n 21

<210> 189
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (20)..(21)
<223> n = deoxythymine

<400> 189
cccaucccgg aagaguucan n 21

<210> 190
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (20)..(21)
<223> n = deoxythymine

<400> 190
ggagaugcag guauuggugn n 21

<210> 191
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature

<222> (20)..(21)
<223> n = deoxythymine

<400> 191
caccaauacc ugcaucuccn n 21

<210> 192
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (20)..(21)
<223> n = deoxythymine

<400> 192
cucuuccggg auggggguaan n 21

<210> 193
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (20)..(21)
<223> n = deoxythymine

<400> 193
uuaccccauc ccggaagagn n 21

<210> 194
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 194
gaucgucgcc accuggaugc u 21

<210> 195
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 195
cauccaggug gcgacgaucu u 21

<210> 196
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 196
ugcauguucg uggccucuaa g 21

<210> 197
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 197
uagaggccac gaacaugcaa g 21

<210> 198
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 198
cuccuacgau acgcuacuau a 21

<210> 199
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 199
uaguagcgua ucguaggagu g 21

<210> 200
<211> 21
<212> RNA
<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 200
uuggucaccu ucuacgagga c 21

<210> 201
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 201
ccucguagaa ggugaccaac u 21

<210> 202
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 202
augauuccug gucaucuggu a 21

<210> 203
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 203
ccagaugacc aggaaucaua g 21

<210> 204
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 204
gacuaugcuu caggcuacca u 21

<210> 205
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 205
gguagccuga agcauaguca u 21

<210> 206
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 206
augauugguc ucgucuggcg c 21

<210> 207
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 207
gccagacgag accaaucauc a 21

<210> 208
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 208
gaguaacaag cucacgcagu u 21

<210> 209
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 209
cugcgugagc uuguuacucg u 21

<210> 210
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 210
ucauacgcgg caggaccaag c 21

<210> 211
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 211
uugguccugc cgcguaugau u 21

<210> 212
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 212
augauagacg cagauagucg c 21

<210> 213
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 213
gacuaucugc gucuaucauc c 21

<210> 214
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 214
gagaagacca cgacauucca a 21

<210> 215
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 215

ggaaugucgu ggucuucucc u 21

<210> 216
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 216
ugaaccuggc caugugcuac c 21

<210> 217
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 217
uagcacaugg ccagguucag a 21

<210> 218
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 218
gacacgcagu ugcucgcuuu u 21

<210> 219
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 219
aagcgagcaa cugcguguca a 21

<210> 220
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 220
gagcugaucu uagauccggc g 21

<210> 221
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 221
ccggaucuaa gaucagcuca a 21

<210> 222
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 222
gacuaaggcu ggucuguguu u 21

<210> 223
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 223
acacagacca gccuuagucu u 21

<210> 224
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 224
gaugugaacg aguccaacgu g 21

<210> 225
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 225
cguuggacuc guucacaucu g 21

<210> 226
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 226
ccuuccgccu cuucaaguac g 21

<210> 227
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 227
uacuugaaga ggcggaaggc a 21

<210> 228
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 228
uccagccuca aguucgguuu u 21

<210> 229
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 229
aaccgaacuu gaggcuggau a 21

<210> 230
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 230
aacugaccgc aaacgaggcc g 21

<210> 231
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 231
uagauccgga aguuacucuc g 21

<210> 232
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 232
uuccugcgag ugaacucguc a 21

<210> 233
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 233
acgaguucac ucgcaggaau c 21

<210> 234
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 234
cagaguaaug cagaaugcgu g 21

<210> 235
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 235
cgcauucugc auuacucuga c 21

<210> 236

<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 236
auccgcuucg acuauggcua c 21

<210> 237
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 237
agccauaguc gaagcggaug a 21

<210> 238
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 238
acugagugug cucuuagcuc u 21

<210> 239
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 239
agcuaagagc acacucaguu c 21

<210> 240
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 240
ucuuccgcuu cucuggacug c 21

<210> 241
<211> 21

<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 241
aguccagaga agcggaagaa g 21

<210> 242
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 242
cuccuaugac gugucugcag g 21

<210> 243
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 243
ugcagacacg ucauaggaga u 21

<210> 244
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 244
aaccgcacug accagcaaug c 21

<210> 245
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 245
auugcugguc agugcgguua g 21

<210> 246
<211> 21
<212> RNA

<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 246
cucaucaucg aggacgacau c 21

<210> 247
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 247
ugucguccuc gaugaugagu u 21

<210> 248
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 248
cauccaggac uguaugugga g 21

<210> 249
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 249
ccacauacag uccuggauga u 21

<210> 250
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 250
aacucuugug cguaaggaaa a 21

<210> 251
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 251
uuccuuacgc acaagaguuc c 21

<210> 252
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 252
acuggcguug aguacguugu u 21

<210> 253
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 253
caacguacuc aacgccaguc a 21

<210> 254
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 254
cacaaggcgu gucaguugau g 21

<210> 255
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 255
ucaacugaca cgccuuguga a 21

<210> 256
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 256
auauaaccga gcagugucuc u 21

<210> 257
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 257
agacacugcu cgguuauaug g 21

<210> 258
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 258
aagcuccucc gcaacaguuu c 21

<210> 259
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 259
aacuguugcg gaggagcuug u 21

<210> 260
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 260
ccaggaacgu guucagccau c 21

<210> 261
<211> 21
<212> RNA
<213> Artificial sequence

<220>

<223> Synthetic oligonucleotide

<400> 261
uggcugaaca cguuccuggu a 21

<210> 262
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 262
ucuguugacg ucuaugccau g 21

<210> 263
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 263
uggcauagac gucaacagag g 21

<210> 264
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 264
accuccgcaa caacuacgcg c 21

<210> 265
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 265
gcguaguugu ugcggaggua g 21

<210> 266
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 266
ccuccgcacc uauaucaugc c 21

<210> 267
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 267
caugauauag gugcggaggu c 21

<210> 268
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 268
aaggaugcuc cucuuccaga u 21

<210> 269
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 269
cuggaagagg agcauccuug u 21

<210> 270
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 270
cuuccagcca uccuaugucc u 21

<210> 271
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 271
gacauaggau ggcuggaaga a 21

<210> 272
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 272
aguacaugga ggacguuacu c 21

<210> 273
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 273
guaacguccu ccauguacuu c 21

<210> 274
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 274
accaugcgac gcagacuuag g 21

<210> 275
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 275
uaagucugcg ucgcauggug a 21

<210> 276
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 276

acugcugagg aacuccuaga g 21

<210> 277
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 277
cuaggaguuc cucagcaguc u 21

<210> 278
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 278
guagaccuug cuacugccug c 21

<210> 279
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 279
aggcaguagc aaggucuacc a 21

<210> 280
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 280
agaucgcaca gcaugaugcu c 21

<210> 281
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 281
gcaucaugcu gugcgaucug u 21

<210> 282
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 282
uacgcuucca gcuacgacug g 21

<210> 283
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 283
agucguagcu ggaagcguau g 21

<210> 284
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 284
gacauagacc uggaccagcc a 21

<210> 285
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 285
gcugguccag gucuauguca a 21

<210> 286
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 286
agcaccugcg uaugaaucua u 21

<210> 287
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 287
agauucauac gcaggugcuc u 21

<210> 288
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 288
caugacggaa cuagagacag c 21

<210> 289
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 289
ugucucuagu uccgucaugg u 21

<210> 290
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 290
ggaccuggac gccaauggag a 21

<210> 291
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 291
uccauuggcg uccagguccu u 21

<210> 292
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 292
uggacauugu agccgcugac g 21

<210> 293
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 293
ucagcggcua caauguccau c 21

<210> 294
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 294
cucuacgcuu guacgaggag c 21

<210> 295
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 295
uccucguaca agcguagaga c 21

<210> 296
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 296
cuauccguac cugcagcacg a 21

<210> 297

<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 297
gugcugcagg uacggauagg u 21

<210> 298
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 298
gaacacuccg cgcuucuuag a 21

<210> 299
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 299
uaagaagcgc ggaguguucu g 21

<210> 300
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 300
acagagcgug gagagcauug c 21

<210> 301
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 301
aaugcucucc acgcucuguu c 21

<210> 302
<211> 21

<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 302
guagcgacag ugacauugaa u 21

<210> 303
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 303
ucaaugucac ugucgcuacu c 21

<210> 304
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 304
cagacuucgg aguaccugcg c 21

<210> 305
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 305
gcagguacuc cgaagucugu u 21

<210> 306
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 306
uaauuggcag cagaucaggu c 21

<210> 307
<211> 21
<212> RNA

<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 307
ccugaucugc ugccaauuac c 21

<210> 308
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 308
guuccauucc gagcuacaca a 21

<210> 309
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 309
guguagcucg gaauggaaca a 21

<210> 310
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 310
uguuaccgca ggacugugug u 21

<210> 311
<211> 21
<212> RNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 311
acacaguccu gcgguaacag u 21

<210> 312
<211> 50
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<220>
<221> misc_feature
<222> (1)..(50)
<223> Portion of enhanced GFP covering region B

<400> 312
cctacggcaa gctgaccctg aagttcatct gcaccaccgg caagctgccc 50

<210> 313
<211> 19
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 313
ggaagtgcag ttagggctg 19

<210> 314
<211> 19
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 314
cagcagcagc aaactggcg 19